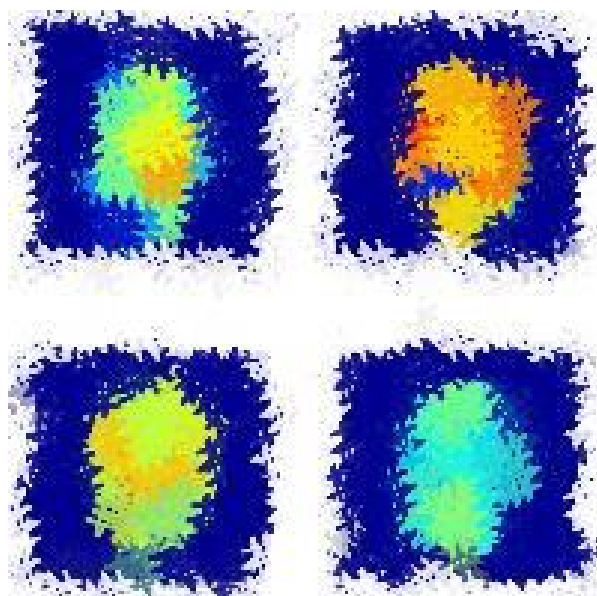


UNIVERSIDAD DE CÓRDOBA
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS AGRÓNOMOS
DEPARTAMENTO DE PRODUCCIÓN ANIMAL



**“APLICACIÓN DE LAS CÁMARAS
HIPERESPECTRALES AL CONTROL DE CALIDAD EN
LA INDUSTRIA AGROALIMENTARIA”**

TESIS DOCTORAL REALIZADA POR
CRISTÓBAL GARRIDO NOVELL

DIRIGIDA POR:

ANA GARRIDO VARO (CATEDRÁTICA DE UNIVERSIDAD)
DOLORES PÉREZ MARÍN (CATEDRÁTICA DE UNIVERSIDAD)

TITULO: *Aplicación de las cámaras hiperespectrales al control de calidad en la industria agroalimentaria*

AUTOR: *Cristóbal Garrido Novell*

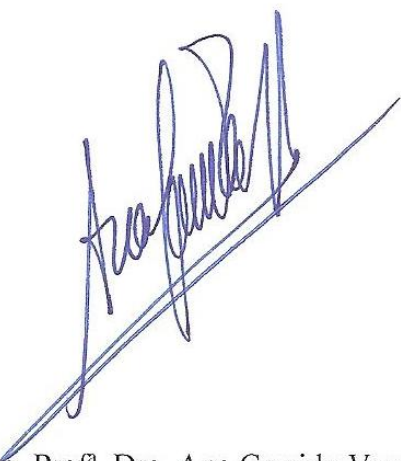
© Edita: UCOPress. 2019
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

<https://www.uco.es/ucopress/index.php/es/>
ucopress@uco.es

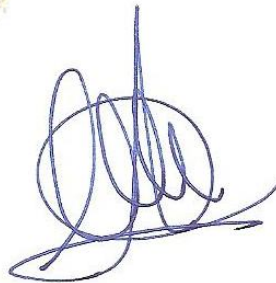
Dra. ANA GARRIDO VARO, Catedrática de Universidad del Departamento de Producción Animal de la Universidad de Córdoba, y Dra. DOLORES PÉREZ MARÍN, Catedrática de Universidad del Departamento de Producción Animal de la Universidad de Córdoba

INFORMAN: que la presente tesis doctoral titulada “Aplicación de cámaras hiperespectrales al control de calidad en la industria agroalimentaria”, y que constituye la memoria presentada por D. Cristóbal Garrido Novell, ha sido realizada bajo nuestra dirección durante el periodo 2011-2018 y, a nuestro parecer, reúne las condiciones académicas exigidas por la legislación vigente para aspirar al grado de Doctor por la Universidad de Córdoba.

Y para que quede constancia, firmamos el presente informe en Córdoba a 11/09/2018



Fdo. Prof^a. Dra. Ana Garrido Varo



Fdo. Prof^a. Dra. Dolores Pérez Marín



TÍTULO DE LA TESIS: Aplicación de las cámaras hiperespectrales al control de calidad en la industria agroalimentaria.

DOCTORANDO/A: Cristóbal Garrido Novell.

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

El doctorando se incorporó en el año 2011, al Grupo de Investigación de Ingeniería de Sistemas de Producción Agro-ganaderos (AGR-128, antes Zootecnia), primero como contratado con cargo a proyectos y después como becario de Formación del Profesorado Universitario.

Su trabajo de Tesis Doctoral, ha conllevado una amplia formación en múltiples aspectos, ligados al análisis NIR-imagen. Dado que el grueso de su trabajo de investigación debería realizarse con un instrumento NIR-imagen se posibilitó una formación específica en dicha instrumentación, sobre todo a través del contacto continuo con especialistas en la materia, especialmente en sus visitas a la Universidad de Córdoba (Dr. Moon S. Kim , USDA-Beltsville, EEUU y Dr. José Manuel Amigo, Universidad de Copenhague).

Tras la optimización del equipo NIR-imagen, no comercial y construido partir de la compra independiente de sus partes, se llevó a cabo la protocolización del análisis y la formación de las bibliotecas de imágenes en manzana, jamón curado y harinas animales. Una vez conseguidas unas bases de datos hiperespectrales de suficiente calidad se llevó a cabo el análisis multivariante de las mismas (aspecto este de gran complejidad, tanto por el elevado volumen de datos manejados, como por la sofisticación de los algoritmos a emplear). Esta actividad se ha simultaneado, con asistencias Congresos y Seminarios, con la redacción y publicación de artículos y con la propia redacción del documento de Tesis.

El doctorando ha tenido múltiples oportunidades de formación y discusión de resultados en Córdoba, durante las estancias y visitas que desde hace 4 años realiza anualmente el Prof. Dr. Tom Fearn, Director del Departamento de Estadística de la University College London, figura muy relevante del tratamiento estadístico y multi e hipervariante de datos espectrales. Del mismo modo la estancia anual del investigador principal del departamento de sensores del United State Department of Agriculture Moon S. Kim, una de las máximas autoridades mundiales en cuanto a equipamiento hiperespectral, ha aportado al doctorando un bagaje considerable en cuanto al manejo y optimización de datos hiperespectrales y su tratamiento.

La Tesis Doctoral realizada por Cristóbal Garrido Novell es de novedad científica, tanto por la instrumentación utilizada, como por la metodología de tratamiento de datos. Supone un importante punto de partida para el desarrollo de aplicaciones comerciales en industrias de nuestro entorno de considerable repercusión socio-económica (cárnica y hortofrutícola). Asimismo avanza y aporta nuevas perspectivas en cuanto a la optimización de la detección de la especie animal en subproductos proteicos de origen animal se refiere, aspecto éste de vital importancia para que la Comisión Europea levante la prohibición de alimentar una especie con harinas procedente de su misma especie, evitando así el canibalismo, y contribuyendo a la minimización del riesgo de EEB.

Publicaciones.

Garrido-Novell, C., Pérez-Marín, D., Amigo, J. M., Fernández-Novales, J. F., Guerrero, J. E., Garrido-Varo, A. 2012. Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras. Journal of Food Engineering 113 (2012) 281-288

Garrido-Novell, C., Garrido-Varo, A., Pérez-Marín, D., Guerrero-Ginel, J.E. 2015. Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging, Journal of Food Engineering 153 (2015) 117-123

Yang, C.C., Garrido-Novell, C., Pérez-Marín, D., Guerrero-Ginel, J.E., Garrido-Varon, A., Cho, H., Kim, M.S. 2015. Differentiation of beef and fish meals in animal feeds

using chemometric analytic models, Journal of Biosystems Engineering, 40(2):153-158. (2015.6)

Garrido-Novell, C., Garrido-Varo, A., Pérez-Marín, D., Guerrero, J.E. 2018. Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins. Chemometrics and Intelligent Laboratory Systems, 172 (2018) 90-99

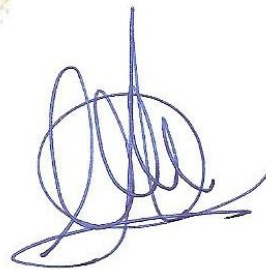
Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 11 de septiembre de 2018.

Firma del/de los director/es



Fdo. Prof^a. Dra. Ana Garrido Varo



Fdo. Prof^a. Dra. Dolores Pérez Marín

Índice

Índice de figuras.....	11
Índice de tablas.....	12
Lista de abreviaturas y símbolos más frecuentes	13
1. INTRODUCCIÓN.....	16
2. OBJETIVOS.....	19
3. REVISIÓN BIBLIOGRÁFICA	21
3.1. Espectroscopía de Reflectancia en el Infrarrojo Cercano y otras técnicas espectroscópicas con aplicación en la imagen hiperespectral.....	21
4.2. Análisis de imagen en teledetección.	25
4.3. Del análisis de imagen y la Espectroscopía a la Imagen Hiperespectral.....	26
4.4. Sistemas de Imagen Hiperespectral.....	27
4.4.1. Adquisición de imágenes hiperespectrales mediante barrido lineal.....	29
4.4.2. Componentes de un sistema de imagen hiperespectral de barrido lineal.....	31
Fuentes de luz.....	31
Espectrógrafos.....	32
Sensores de imagen.....	33
4.2. Herramientas matemáticas para el análisis de imágenes hiperespectrales.	34
4.2.1. Pretratamiento espectral.....	36
4.2.2. Exploración del hipercubo. Análisis por Componentes Principales y detección de anómalos.....	38
4.2.3. Predicción.....	42
Partial Least Squares (PLS).....	42
Multivariate Curve Resolution (MCR) y Classical Least Squares (CLS).....	44
4.2.4. Segmentación y/o clasificación.....	46
Partial Least Squares Discriminant Analysis (PLS-DA)	46
Linear Discriminant Analysis (LDA)	47
K-MEANS y Fuzzy Cluster Means Analysis (FCM).....	48
4.2.5. Análisis textural	50
Matriz de coocurrencia	50
Fractales	52
4.2.6. Árboles de regresión.....	53
4.3. Aplicaciones.....	55

4.3.1. Frutas y hortalizas.	55
4.3.2. Productos de origen animal.	61
4.4. Bibliografía.....	67
5. RESULTADOS Y DISCUSIÓN.....	84
5.1.-Puesta a punto del equipo hiperspectral.	84
5.1.1. Metodología.	87
Señal/Ruido (SNR).....	87
Nitidez spectral.....	90
5.1.2. Resultados de la comparación.	90
Señal/Ruido (SNR).....	90
5.1.3. Conclusiones de la comparación.	97
5.2. “Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, <i>Journal of Food Engineering</i> 113 (2012) 281-288.....	99
5.3. “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”, <i>Journal of Food Engineering</i> 153 (2015) 117-123 y “Methodology for marbling characterisation of pure Iberian cured loin based on hyperspectral imaging”. <i>Comunicación oral 16th International Conference on Near Infrared Spectroscopy</i> . 2013.	109
5.4. “Differentiation of beef and fish meals in animal feeds using chemometric analytic models”, <i>Journal of Biosystems Engineering</i> , 40(2):153-158. (2015.6)	122
5.5. “Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, <i>Chemometrics and Intelligent Laboratory Systems</i> , 172 (2018) 90-99.....	129
6. CONCLUSIONES.....	139

Índice de figuras.

Figura 1 Fenómenos de interacción entre la radiación electromagnética y la materia (Zamora, 2013).	23
Figura 2. Croquis de los sistemas de adquisición de imágenes hiperespectrales. (Amigo et al. 2010).	29
Figura 3. Esquema del análisis de una imagen hiperespectral.....	35
Figura 4. Imagen, espectros NIRS y “scores” tras Análisis de Componentes Principales de una patata contaminada con estiércol y tierra (elaboración propia).	39
Figura 5. Rotación de los ejes de referencia tanto en la variable dependiente como en la independiente en el sentido de la máxima varianza (Fuente: models.life.ku.dk).	42
Figura 6. Imagen de la clasificación realizada por el FCM, espectros centrales de las clases encontradas y matrices de coocurrencia de dichas imágenes en rodajas de caña de lomo.	50
Figura 7. Equipo de barrido lineal modular no comercial.....	85
Figura 8. Equipo Matrix NIR integral de laboratorio, de barrido espectral, con las dos lámparas halógenas de 250 w.	86
Figura 9. Posiciones de las cerámicas de referencia en el equipo de barrido espectral.....	89
Figura 10. Posiciones de las cerámicas de referencia en el equipo de barrido lineal.	90
Figura 11. Señal con todos los píxeles (antes de la selección aleatoria de píxeles) para el equipo de barrido lineal.....	91
Figura 12. Señal con todos los píxeles (antes de la selección aleatoria de píxeles) para el equipo de barrido espectral.	92
Figura 13. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la máxima resolución (1x1) en el sistema de barrido lineal.	93
Figura 14. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 2x2 en el sistema de barrido lineal.....	93
Figura 15. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 5x5 en el sistema de barrido lineal.....	94
Figura 16. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 10x10 en el sistema de barrido lineal.....	94
Figura 17. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la máxima resolución (1x1) en el sistema de barrido espectral.....	95
Figura 18. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 2x2 en el sistema de barrido espectral.	95
Figura 19. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 5x5 en el sistema de barrido lineal.....	96
Figura 20. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 10x10 en el sistema de barrido lineal.....	96
Figura 21. Nitidez espectral (Full Width at Half Maximum) para el equipo de barrido espectral y el de barrido lineal.....	97

Índice de tablas.

Tabla 1. Aplicaciones de equipos hiperspectrales en frutas y hortalizas. Se indica el producto, el objeto del análisis, el rango espectral, la precisión de la predicción o clasificación y la referencia.....	57
Tabla 2. Aplicaciones de equipos hiperspectrales en productos de origen animal. Se indica el producto, el objeto del análisis, el rango espectral, la precisión de la predicción o clasificación y la referencia.	64
Tabla 3. Resumen de las réplicas tomadas para cada uno de los equipos.	89

Lista de abreviaturas y símbolos más frecuentes

%: porcentaje

2D: dos dimensiones

3D: tres dimensiones

A/D: transformación de analógico en digital

ALS: alternantes, Alternating

ANOVA: análisis de la varianaza, Analysis Of Variance

BIL: bandas intercaladas por líneas, Band Interleaved by Line

BSQ Secuencia de bandas, Band Sequential

C: contraste

CCD: dispositivo de carga acoplada, Charge Coupled Device

C-H: enlace carbono hidrógeno

CLS: mínimos cuadrados clásicos, Clasical Least Squares

CMOS semiconductores de óxido metálico complementarios, complementary metal-oxide-semiconductor

C-N: enlace carbono nitrógeno

C-O: enlace carbono oxígeno

CR: correlación

DA: análisis discriminante, Discriminant Analysis

DT: Detrend

E: energía

EMCCD: dispositivos de carga acoplada multiplicadores de electrones, Electron-Multiplying Charge Coupled Device

FCM: agrupamiento difuso, Fuzzy Cluster Means

FOV: campo de vision, Field of View

FWHM: anchura completa a la mitad del máximo, Full Width at Half Maximum

GH: Distancia de Mahalanobis global, Global Mahalanobis Distance

GLM: matriz de coocurrencia, Gray-level cooccurrence matrix

H: homogeneidad

I+ D + i: Investigación, Desarrollo e innovación

Inf: infinito

-Inf: menos infinito

LDA: análisis discriminante lineal, Linear Discriminant Analysis

LED: diodos de emisión de luz, Light-Emitting Diodes

LV: variable latente, Latent Variable

MCR: resolución curva multivariante, Multivariate Curve Resolution

MCT: mercurio-cadmio-telurio

MSC: corrección multiplicativa de la dispersión, Multiplicative Scatter Correction

MWIR: infrarrojo medio

NaN: resultado no numérico, Not a Number

NH: Distancia de Mahalanobis de vecinos, Neighbourhood Mahalanobis Distance

NIRS: Espectroscopía del Infrarrojo Cercano, Near Infrared Spectroscopy,

nm: nanómetro

PC: componente principal (principal component).

PCA: Análisis de componentes principales, Principal Component Analysis

PGP: prisma de dispersión, prism-grating prism

PLS: Mínimos cuadrados parciales, Partial Least Squares

PLS: mínimos cuadrados parciales, Partial Least Squares

QE: eficiencia cuántica, Quantum Efficiency

QTH: cuarzo-tungsteno

R²: coeficiente de determinación

RGB: Red, Green and Blue; composición del color en base a estos tres colores primarios

RMSEP: error cuadrático medio de predicción, Root Mean Standard Error of Prediction

RMSEPG: error cuadrático medio de predicción global, Root Global Mean Standard Error of Prediction

ROI: region de interés, Region of Interest

SEC: error estándar de calibración, Standard Error of Calibration

SECV: error estándar de validación cruzada (Standard Error of Cross Validation)

SG: desviación estándar global, Standard Global deviation

SNR: ratio señal-ruido, Signal to Noise

SNV Standard Normal Variate

USDA: Departamento de Agricultura de los Estados Unidos, United State Department of Agriculture.

UV: ultravioleta

VIP: proyección importante de variables, Variables Important Projection

VIS: visible

VNIR: visible e infrarrojo cercano

λ : longitud de onda

1. *INTRODUCCIÓN*

En la actualidad, se ha incrementado el interés por los productos agroalimentarios de máxima calidad, sobre todo en aquellos productos con alto valor añadido y, asimismo, se requiere cada día más información sobre la trazabilidad y seguridad de los productos por parte del consumidor. Así, pequeñas reducciones de calidad pueden tener un efecto muy duradero y perjudicial sobre un producto o marca comercial. La necesidad de llevar a cabo un control riguroso y fiable que incluya aspectos relacionados con la trazabilidad, calidad y seguridad de los productos agroalimentarios es una realidad. Esto ha propiciado que todo lo referente a ensayos, análisis, tecnologías y mediciones rápidas, precisas, exactas y de bajo coste, sea de vital importancia para la industria.

Una industria alimentaria con visión de futuro, obviamente, debe continuar desarrollando su tecnología de producción para poder producir alimentos que cumplan tanto con la legislación actual como con las expectativas y demandas de los consumidores. Garantizar productos de alta calidad y salubridad, sin comprometer al mismo tiempo la competitividad, optimizando los costos del proceso productivo y su control, son aspectos esenciales en el actual contexto europeo e internacional.

Las nuevas tecnologías de producción en sí mismas no crean ni garantizan una alta calidad de los productos ni la optimización de los procesos de producción. Esto requiere conocimiento sobre las nuevas oportunidades tecnológicas para actuar y controlar la calidad en diferentes puntos de la cadena agroalimentaria. Si bien es cierto que los métodos de análisis tradicionales (químicos, físicos, sensoriales, etc.) están en continua renovación, y que es generalmente aceptado el hecho de que los laboratorios son capaces de determinar cualquier constituyente de un alimento, es igualmente reconocido que la alta sofisticación del equipamiento, junto con la lentitud y el alto coste por muestra de los métodos tradicionales, dificulta su implantación en los programas actuales de control y gestión de calidad a nivel de producción.

Son numerosas las tecnologías que se están aplicando para el uso de información y toma de decisiones “en tiempo real”. La Espectroscopía de Reflectancia en el Infrarrojo Cercano (NIRS) es una tecnología ideal para satisfacer las demandas actuales de

productores y consumidores ya que, entre otras características, es una tecnología rápida, precisa, no destructiva, no contaminante, fácil de utilizar, muy versátil y económica. Desde la década de los 70, numerosas publicaciones científicas avalan la capacidad e idoneidad de la tecnología NIRS para su uso en programas de trazabilidad, aseguramiento de la calidad y etiquetado en el sector agroalimentario (Garrido y Pedro de, 2007). El Grupo de Ingeniería de Sistemas de Producción Agro-Ganaderos (AGR-128), en el marco del cual se ha desarrollado la presente Tesis Doctoral, tiene mas de 30 años de experiencia en I+ D +i en el terreno de la tecnología NIRS. Mas recientemente, conscientes de la importancia de seguir avanzando en la complementación de sensores para su uso en el sector agroalimentario, dicho grupo ha iniciado proyectos de I + D + i en el terreno de la combinación de las tecnologías NIRS y de imagen. La tecnología de imagen espectral combina las fortalezas de la imagen convencional con las de la espectroscopía, de forma tal que es posible abordar aplicaciones a las que ambas tecnologías por separado no podrían responder. Como indica McClure (2003), la tecnología de imagen hiperespectral está en los inicios de su desarrollo y ocupará la escena analítica del siglo XXI.

En la última década, ha existido un gran impulso para la puesta en el mercado de cámaras hiperespectrales que operan en el rango del visible y del infrarrojo cercano. Sin embargo, el desarrollo del conocimiento científico en este terreno ha estado limitado por numerosos factores, tales como el limitado rango de las cámaras para cubrir el rango completo en la región NIR (780-2500 nm), niveles de ruido elevados, la necesidad de uso de ordenadores de gran capacidad de computación, el desarrollo de algoritmos matemáticos adecuados para procesar la enorme cantidad de datos generados, la ausencia de conocimiento, el elevado precio de los instrumentos, etc. (Geladi et al., 2004; Buttingsrud y Alsberg, 2006).

En el momento actual, gran parte de estos problemas han sido resueltos y la combinación NIRS-Imagen permite obtener un espectro NIR por cada píxel, generando unas imágenes que posibilitan explorar la interdependencia de información espectral y espacial de una muestra. Esto es, permite realizar una medida precisa en el terreno espectral, espacial y temporal, y la interdependencia de los tres tipos de medidas nos permite cuantificar propiedades importantes. Diferentes estudios confirman que el control en tiempo real con tecnología NIRS-imagen será imprescindible en los sistemas modernos

de control industrial y clasificación en línea en el futuro cercano (Gowen et al, 2007; ElMasry et al., 2012, Wu y Da-Wen Sun, 2013).

Sin embargo, son numerosos los aspectos a resolver para que dichos sensores se incorporen de una forma cotidiana al control de sistemas agrarios complejos. Es necesario seguir avanzando en el desarrollo y optimización instrumental y, muy especialmente, en la integración de la información generada en distintas plataformas, así como en el desarrollo de protocolos para la automatización de la captura y tratamiento de la señal.

La presente Tesis Doctoral trata de contribuir a la generación de conocimiento en este ámbito, para lo cual se han fijado los objetivos que se mencionan a continuación.

2. OBJETIVOS

Objetivo General:

Desarrollar metodologías de apoyo a sistemas de control de procesos en industrias agro-alimentarias de nuestro entorno (harinas cárnicas, cerdo ibérico y frutas) mediante la combinación de sensores de imagen y espectroscopía de infrarrojo cercano.

Objetivos Específicos:

1. Evaluar el potencial de los sensores hiperespectrales para la predicción y/o categorización de parámetros de calidad relevantes de frutas, al objeto de determinar la vida útil de las mismas [La consecución del objetivo queda reflejada en la publicación: “Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288)].

2. Evaluar el potencial de los sensores hiperespectrales en la industria del jamón ibérico para la categorización de sus productos por tipo de alimentación y contenido en sal, a la vez que se estudia la distribución de la misma en los distintos músculos del jamón tras su curado [La consecución del objetivo queda reflejada en la publicación: “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”, *Journal of Food Engineering* 153 (2015) 117-123)].

3. Evaluar el potencial de los sensores hiperespectrales para la cuantificación y detección de contaminaciones de harinas de rumiantes en harinas de carne [La consecución del objetivo queda reflejada en la publicación: “Differentiation of beef and fish meals in animal feeds using chemometric analytic models”, *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)].

4. Evaluar el potencial de los sensores hiperespectrales y realizar propuestas de optimización para la identificación de la especie animal en harinas de carne [La consecución del objetivo queda reflejada en la publicación: “Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99].

3. REVISIÓN BIBLIOGRÁFICA

Antes de la presentación de los antecedentes, la metodología y los resultados de cada uno de los objetivos a través de las publicaciones científicas generadas, se realizará una breve revisión de los fundamentos de la tecnología de imagen hiperespectral, partiendo de una breve reseña de sus antecesoras: la Espectroscopía del Infrarrojo Cercano (NIRS) y la Visión Artificial. Posteriormente, se expondrán los principios, la instrumentación y los métodos de análisis de los datos en imagen hiperespectral, haciendo especial hincapié en el equipo usado para la consecución de los cuatro objetivos de la presente tesis. Finalizaremos la introducción con una revisión de las aplicaciones de la imagen hiperespectral en el ámbito agroalimentario.

3.1. Espectroscopía de Reflectancia en el Infrarrojo Cercano y otras técnicas espectroscópicas con aplicación en la imagen hiperespectral.

Descubierta por Herschel en 1800 (Herschel, 1800), la radiación en el infrarrojo cercano fue desarrollada tecnológicamente e instrumentalmente por Karl Norris en el Departamento de Agricultura de los Estados Unidos (USDA, United State Department of Agriculture) (Hart y Norris, 1962). La Espectroscopía de Reflectancia del Infrarrojo Cercano (NIRS, por sus siglas en inglés de Near Infrared Reflectance Spectroscopy) es una técnica espectroscópica vibracional basada en la región espectral del infrarrojo cercano (de los 780 a los 2.500 nm).

Cuando la radiación NIRS interactúa con la materia puede ser absorbida, reflejada difusamente (cuando parte de la luz incidente que llega a la superficie penetra en la muestra y es reflejada conteniendo información) o transmitirse, cuando parte de la radiación atraviesa la muestra completamente. También ocurren fenómenos de pérdida de energía por refracción, reflexión especular o dispersión, éste último conocido como “efecto scatter” (dispersión de la luz cuando la radiación no reflejada que penetra la muestra es dispersada en múltiples direcciones), viéndose afectada por el tamaño de partícula y el índice de refracción (Shenk y Westerhaus, 1991; Dahm y Dahm, 2001.; Siesler et al., 2002).

En la región del infrarrojo medio, las vibraciones moleculares son el origen de las bandas de absorción fundamentales, mientras que en el infrarrojo cercano se localizan bandas de absorción que se corresponden con sobretonos y combinaciones de las bandas fundamentales (Miller, 2001; Benson, 2003.). Estas absorciones se producen al vibrar, plegarse, alargarse o deformarse las moléculas (Lachenal, 2000.). Para que exista absorción de la luz incidente, es necesario que la molécula experimente un cambio de su momento dipolar como consecuencia de su vibración (Siesler et al., 2002).

Las moléculas obedecen a las leyes de la mecánica cuántica, de forma que la energía es cuantificada en ciertos niveles. Dado que las moléculas no se comportan como osciladores armónicos perfectos, sino que presentan distintos grados de no armonicidad, las absorciones NIRS tienen lugar como sobretonos de absorciones fundamentales, a diferencia de lo que ocurre en el infrarrojo medio; si bien los niveles de energía continúan obedeciendo las leyes de la mecánica cuántica. Otra fuente de bandas de absorción NIR son las combinaciones que tienen lugar cuando una molécula comparte energía NIR de un fotón entre dos o más absorciones fundamentales. Las bandas de combinación más conocidas ocurren con el alargamiento y el plegado de un mismo enlace. Sin embargo, existen otras cuya posición e intensidad resultan difíciles de predecir (Ciurezak, 2001). Pese a que éstas no pueden ser interpretadas o asignadas a un compuesto químico determinado, contienen información sobre los enlaces C-H, C-O y C-N (Shenk et al., 1992).

La facultad clasificadora del espectro NIRS se basa en su natutaleza única para cada muestra, no en vano se conoce por huella espectral. Su capacidad para cuantificar diferentes componentes se basa en que los cambios en la respuesta del espectro NIRS son proporcionales a los cambios en la concentración de los componentes químicos que los provocaron, así como a ciertas características físicas de la muestra (Workman y Shenk, 2004).

De los distintos fenómenos que pueden producirse cuando un haz de luz NIRS interactúa con una muestra, se derivan distintos modos de análisis cuando se trabaja con la Espectroscopía NIRS, y por extensión con dispositivos de imagen hiperespectral. Los dos modos de análisis principales son la reflectancia y la transmitancia (Fig. 1).

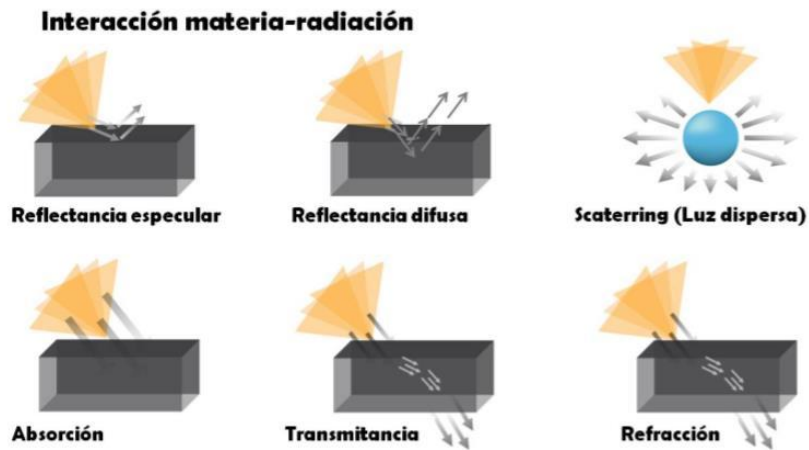


Figura 1 Fenómenos de interacción entre la radiación electromagnética y la materia (Zamora, 2013).

Cuando una muestra biológica es expuesta a una fuente de iluminación (p.ej. luz halógena), una pequeña cantidad de la luz incidente es reflejada provocando reflectancia especular. El resto de energía incidente penetra en la muestra, sufriendo los fotones múltiples dispersiones (“scatter”) antes de ser absorbidos o reflejados para reemerger desde la superficie de la muestra, generando la reflectancia difusa. Así, las medidas de reflectancia, generalmente, recogen señal de reflectancia especular y difusa. Cuando la muestra es pequeña y la iluminación es lo suficientemente fuerte. La luz incidente puede también provocar transmitancia difusa a través de la muestra. Las señales de reflectancia y la transmitancia son normalmente adquiridas en función de la longitud de onda y representadas en forma de espectros. Las técnicas de espectroscopía de transmitancia y reflectancia pueden trabajar en un amplio rango de longitudes de onda desde la ultravioleta hasta la infrarroja. En estos tipos de espectroscopía, la constitución espectral de la luz incidente no se altera con las interacciones con la muestra. La información sobre la muestra se obtiene gracias a los cambios en la intensidad de la luz debido a su reflexión, absorción y dispersión a las distintas longitudes de onda. La espectroscopía de reflectancia suele usarse para la inspección de la superficie de muestras, mientras que la espectroscopía de transmitancia es útil para evaluar los atributos internos de las mismas (Qing y Zude, 2007).

Otras técnicas espectroscópicas que suelen usarse para complementar la Espectroscopía de Infrarrojo Cercano, y que se aplican también en imagen hiperespectral, son la espectroscopía de Fluorescencia y la Espectroscopía Raman. En la Espectroscopía de

Fluorescencia, el fenómeno físico fundamental es la emisión de luz por un compuesto cuando éste es excitado por una radiación incidente. En general, la longitud de onda de la excitación es fija, usándose bandas estrechas de alta intensidad (p. ej. la radiación ultravioleta) como fuentes de excitación fluorescente. La luz emitida, que normalmente tiene una menor energía que la absorbida, es observada normalmente en un rango espectral de mayores longitudes de onda. Este espectro contiene información sobre la composición de la muestra. Las señales de fluorescencia son típicamente presentadas como espectros de emisión, que son una representación de la intensidad *versus* las longitudes de onda. La intensidad de fluorescencia es generalmente proporcional a la concentración de los fluoróforos en la muestra. Muchos tejidos animales y vegetales (p. ej., grasas, hojas y frutas) emiten fluorescencia en la región visible del espectro electromagnético cuando son excitados con radiación ultravioleta. Las emisiones de fluorescencia desde materiales biológicos aparecen típicamente en las regiones del azul, el verde, el rojo y el rojo lejano del espectro. La Espectroscopía de Fluorescencia es considerada una técnica muy sensible de detección, que puede medir pequeños cambios en las muestras biológicas (Jianwei Qin, 2017).

En la Espectroscopía Raman, un rayo de luz monocromático (p. ej. un láser) es absorbido, reflejado, y dispersado cuando sus fotones interaccionan con las moléculas de una muestra biológica. La luz dispersada proviene tanto de dispersiones elásticas como inelásticas de los fotones. La inmensa mayoría de la luz dispersada proviene de dispersiones elásticas (dispersión de Rayleigh), que tiene la misma frecuencia que la luz incidente. Una muy pequeña proporción de los fotones cederán energía a las moléculas de la muestra, llevándolas de un estado basal a un estado excitado (dispersión de Stokes) o la recibirán en el proceso opuesto (dispersión anti-Stokes). Ambos tipos de dispersiones, de Stokes y anti-Stokes, son llamadas dispersión de Raman.

Normamente las medidas Raman registran las dispersiones de Stokes, debido a la mayor intensidad de éstas con respecto a las anti-Stokes. La intensidad de la dispersión de Raman es muy débil (del orden de 1 fotón sobre 10^6 - 10^8 fotones dispersados). Su valor es proporcional a la intensidad del haz de luz incidente y al inverso de la cuarta potencia de la longitud de onda de excitación. Un espectro Raman típico presenta una serie de picos estrechos y puntiagudos. Pueden usarse para determinar la composición de muestras biológicas. La intensidad de los picos Raman es linealmente proporcional a la

concentración de la molécula excitada y puede usarse, por tanto, con fines cuantitativos (Jianwei Qin, 2017)

La construcción de equipos de medida hiperespectral basados en estos tipos de Espectroscopía (reflectancia, transmitancia, fluorescencia y Raman) requiere de dispositivos específicos y de disposiciones de los mismos también específicas. Así, por ejemplo, la Espectroscopía de Transmitancia, requiere de fuentes de luz de mucha intensidad y situadas en la cara opuesta de la muestra de donde se sitúa el sensor (espectrofotómetro y cámara hiperespectral). La Espectroscopía de Fluorescencia requiere de espectrofotómetros en el rango del visible, provistos de filtros que eliminen cualquier interferencia en la medida producida por la fuente de luz ultravioleta; y la Espectroscopía Raman necesita fuentes de luz de muy alta intensidad (láser).

4.2. Análisis de imagen en teledetección.

Estos sistemas empezaron a usarse en teledetección, extendiéndose su uso rápidamente a otros campos como la astronomía, la agricultura, la medicina, la industria farmacéutica y la industria agroalimentaria.

La imagen multiespectral ha sido usada como fuente de información de la tierra y el agua para diversos propósitos desde la década de los 60 (Landgrebe, 1999). Los sistemas multiespectrales suelen recoger los datos en un rango que oscila de tres a seis bandas para cada observación, desde el visible hasta el infrarrojo. El espectro obtenido proviene de la energía reflejada y emitida por la tierra. La obtención de espectros brutos de mayor calidad fue el primer obstáculo que tuvo que superar esta tecnología. En las últimas dos décadas, los sistemas multiespectrales han dado paso a los hiperespectrales, haciendo posible la recogida de cientos de bandas espectrales en cada adquisición de imagen. En esta segunda etapa, los retos del desarrollo tecnológico de la teledetección hiperespectral han ido asociados al coste de los sensores, el volumen de los datos y el tiempo de procesado de los mismos (Govender, 2007).

El mayor peso de la georreferenciación frente a la quimiometría en las aplicaciones de la imagen hiperespectral en teledetección es una valiosa fuente para el desarrollo de la

imagen hiperespectral en aplicaciones agroindustriales. La incorporación a nuestro ámbito de problemas amplia y profundamente estudiados en teledetección, como pueden ser la resolución de mezclas, la caracterización de la distribución de una propiedad en el espacio o el reconocimiento de patrones en las imágenes, es de gran interés. El uso de la teoría de los fractales o la matriz de coocurrencia son algunas de las soluciones aportadas a estos problemas desde el ámbito de la teledetección. En el apartado correspondiente a los modelos matemáticos, se profundizará en las mismas (fractales y matriz de coocurrencia).

4.3. Del análisis de imagen y la Espectroscopía a la Imagen Hiperespectral.

Las técnicas ópticas no invasivas tales como la Espectroscopía y el análisis de imagen se han mostrado extremadamente ventajosas para la inspección en línea de productos agroalimentarios, con el fin de garantizar su calidad y seguridad. Se ha demostrado la valía de la Espectroscopía para la determinación de las cualidades esenciales de los alimentos en base a la medida de sus propiedades ópticas (Bock y Connelly. 2008). Sin embargo, la Espectroscopía no nos da información sobre la distribución espacial de las características de los productos agroalimentarios, aspecto de gran interés en determinadas aplicaciones.

Por otro lado, la medida de características externas de los productos puede alcanzarse por imagen convencional o visión artificial. Sin embargo, la imagen convencional es incapaz de distinguir elementos con color similar, analizar regiones complejas, predecir componentes químicos o detectar defectos invisibles.

Con la integración de las principales ventajas de la Espectroscopía y la imagen, la Imagen Hiperespectral puede simultáneamente adquirir información espectral y espacial. La Imagen Hiperespectral puede ser aplicada a la predicción cuantitativa de propiedades químicas y físicas, así como a su distribución espacial simultáneamente. Si un espectrómetro convencional puede dar respuesta a la pregunta del qué, y la imagen convencional a la pregunta del dónde, la imagen hiperespectral puede dar respuesta a la pregunta del “dónde está el qué”, o visto a nivel de imagen o mapa, del “cómo se distribuye el qué”. Dando un paso más, la imagen hiperespectral puede, sin necesidad de

pasar por la definición de un qué, es decir, de la determinación de sustancias químicas aislables del resto de la imagen, mostrar singularidades en las características de imágenes a longitudes de onda invisibles al ojo humano. Esta capacidad podría ayudar a la clasificación de muestras gracias a diferencias entre ellas puramente espectro-espaciales, que integrarían las diferencias químicas sin explicitarlas.

4.4. Sistemas de Imagen Hiperespectral.

Un sistema de Imagen Hiperespectral produce un cubo de imágenes del mismo objeto a diferentes longitudes de onda. Espectralmente hablando hay dos clases principales: Imagen Multiespectral e Imagen Hiperespectral. El concepto que explica ambos tipos es el mismo. La diferencia entre ellos es el número de imágenes adquiridas para formar el cubo espectral. La Imagen Multiespectral no puede mostrar un espectro real para cada píxel de la imagen. Tiene normalmente menos de diez espectros por banda.

La resolución de los equipos de Imagen Multiespectral suele ser mayor de 10 nm. Además de usar filtros de paso de banda, suelen usar cámaras 3CCD (Dispositivo de Carga Acoplada, Charge-Coupled Device en inglés). Los CCD son circuitos integrados con condensadores acoplados, que pueden transferir su carga a los condensadores vecinos y que convierten la intensidad de luz (fotones) en una señal, primero eléctrica y después digital, gracias a las propiedades fotoeléctricas de ciertos materiales (Janesick , 2001). Las cámaras multiespectrales tienen tres sensores CCD y un prisma divisor del haz dicróico que divide la luz en tres bandas espectrales. Su menor resolución espectral le permite mayores velocidades con respecto a la Imagen Hiperespectral. Mientras que la velocidad de adquisición de las cámaras 3CCD es de docenas de imágenes por segundo, la de las cámaras usadas para la adquisición de Imagen Hiperespectral es de segundos por cada imagen.

El concepto de Imagen Hiperespectral fue introducido en el campo agroalimentario por científicos del Departamento de Agricultura de los Estados Unidos a finales de la década de los 90 (Lu, y Chen, 1999). Su aplicación práctica para la inspección de productos agrolimentarios, en su forma de barrido lineal para reflectancia y

fluorescencia, llegó a principios del nuevo milenio (Kim et al., 2001). Desde entonces la Imagen Hiperespectral ha atraído y sigue atrayendo la atención tanto académica como industrial, y su ámbito de trabajo ha ido expandándose a más sectores y aplicaciones.

Las imágenes hiperespectrales son tridimensionales, con dos dimensiones espaciales y una dimensión espectral. Pueden adquirirse por medio de tres sistemas de adquisición principalmente: por barrido puntual, barrido lineal y barrido superficial (del inglés point-scan, line-scan y area-scan). Los sistemas de barrido lineal, como el usado en la presente tesis, se ajustan mejor al sistema más común de transporte de los productos en la industria agroalimentaria, esto es, las cintas transportadoras (Fig. 2).

En los sistemas de barrido puntual (también llamados en inglés, “whiskbroom method”), un solo espectro es escaneado para cada píxel, moviéndose relativamente el sensor y la muestra en las dos dimensiones espaciales para recoger el resto de espectros. El hipercubo obtenido mediante la integración de todos estos vectores o espectros es almacenado en formato BIP (bandas apiladas por píxel, “band-interleaved-by pixel” en inglés). Este formato es óptimo para el acceso a la información espectral de cada píxel. Las desventajas principales de este sistema son su lentitud a la hora de posicionar la muestra y la necesidad de hardware avanzado para asegurar un posicionado repetible de la muestra.

Los sistemas de barrido superficial (Fig. 2), también conocidos como secuencial de bandas o de barrido espectral son sistemas que en cada secuencia mantienen el campo de imagen fijo y adquieren imágenes monocromáticas en dos dimensiones con toda la información espacial para cada longitud de onda. Esta medida se repite a lo largo de todo el rango espectral, obteniéndose imágenes de todas las bandas que son apiladas en formato BSQ (secuencia de bandas, “band sequential”, en inglés). Este método ofrece un fácil acceso a la información espacial para cada longitud de onda. Como el detector es expuesto a sólo una longitud de onda cada vez puede establecerse un tiempo de exposición adecuado para cada una de ellas. Además, no es necesario que exista un desplazamiento relativo entre la muestra y el sensor, lo cual puede ser útil para ciertas aplicaciones, como las realizadas en laboratorio o las fluorescentes, en las que la muestra ha de permanecer quieta por un tiempo para su excitación. La principal desventaja de este método es su difícil adaptación a la inspección en línea en ambientes industriales, en los que las muestras están en continuo movimiento.

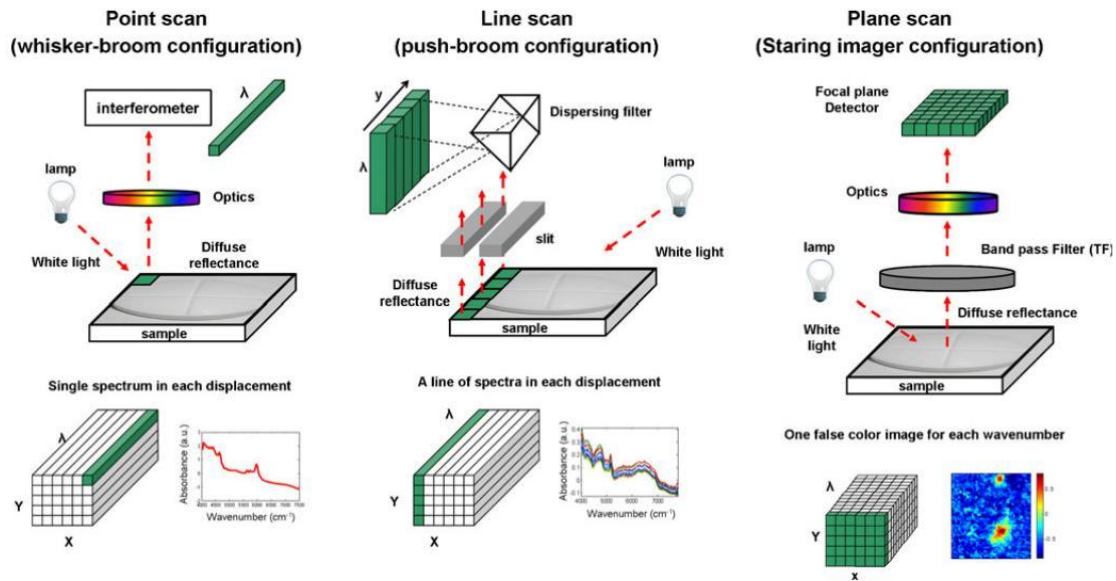


Figura 2. Croquis de los sistemas de adquisición de imágenes hiperespectrales. (Amigo et al. 2010).

4.4.1. Adquisición de imágenes hiperespectrales mediante barrido lineal.

Los sistemas de barrido lineal pueden disponerse de diferentes maneras para satisfacer distintos requerimientos. Dos disposiciones básicas son las de reflexión y transmisión. En modo reflexión, una fuente y un detector son colocados en el mismo lado de la muestra. El detector recoge las señales reflejadas por la muestra: reflectancia especular y difusa, fluorescencia y dispersión trasera (“*backscatter*”) Raman. Todas estas disposiciones son para la inspección superficial (Jianwei, 2017).

En modo transmisión, la fuente de luz y el detector son dispuestos en lados opuestos de la muestra. El detector adquiere las señales que han atravesado la muestra y que pueden ser usadas para evaluar la composición interna de la muestra, especialmente, para muestras delgadas y con unas condiciones internas de baja absorción. Tanto para configuraciones de reflexión como de transmisión, el sistema de barrido lineal adquiere simultáneamente una ranura de información espacial y la información espectral completa de cada uno de los puntos de esa ranura, línea o campo de visión (FOV, de “*field of view*”, en inglés), iluminado o excitado por fuentes lumínicas lineales o superficiales. Por tanto, con el avance de la muestra con respecto a la cámara o de la cámara con respecto a la muestra, se van tomando imágenes de dos dimensiones: una

dimensión espacial y otra espectral. Se toma una imagen con cada avance, paso o “scan”. El hipercubo o imagen espectral en 3D se construye al integrar todos los “scan” adquiridos en la dirección del avance (Jianwei, 2017).

La transmitancia hiperespectral puede obtener información interna de la muestra. Sin embargo, no puede diferenciar la información entre capas individuales de la misma debido a la ausencia de variaciones entre las señales de transmitancia. La Espectroscopía de Resolución Espacial es una técnica que puede obtener información subsuperficial de una muestra laminada. El principio en el que se basa es en separar una fuente de luz puntual y un detector para permitir a la luz viajar a través de un área de la muestra cada vez más profunda. Las señales son obtenidas de una serie de posiciones distintas desplazadas lateralmente desde el punto emisor de luz, permitiendo la evaluación de las capas individuales de la muestra. El concepto ha sido adoptado en cada uno de los modos de adquisición de imagen hiperespectral antes mencionados: reflectancia difusa espacialmente resuelta, fluorescencia difusa espacialmente resuelta y Espectroscopía Raman espacialmente resuelta. Su medida, usando detectores puntuales, es lenta debido a los desplazamientos mecánicos necesarios para la medida de cada punto. Las técnicas basadas en barrido lineal ofrecen un método flexible y eficiente para tomar este tipo de medidas. La imagen de dispersión adquirida de un solo “scan” es un conjunto completo de espectros espacialmente resueltos con la suficiente información espacial y espectral para la evaluación subsuperficial de muestras biológicas heterogéneas y/o laminadas (Jianwei, 2017).

Aunque los sistemas de barrido lineal y los de barrido espectral pueden aplicarse indistintamente en la mayoría de las aplicaciones (existen en el mercado equipos con ambos sistemas de adquisición que cubren desde el visible al infrarrojo), los primeros parecen ser los que más naturalmente se ajustan a las aplicaciones en la industria agroalimentaria. En efecto, la adquisición de imágenes completas, en dos dimensiones, requiere el barrido de la muestra a registrar; barrido para el cual resultan idóneas las cintas transportadoras tan profusamente usadas en la industria agroalimentaria. Sin embargo, la conclusión no es tan inmediata ya que hay grandes diferencias entre ambas tecnologías en cuanto a sus ópticas, su estado de desarrollo, la posibilidad de desarrollar aplicaciones de bajo coste con una u otra, etc.

El diseño de sistemas de imagen hiperespectral portátiles y de bajo coste, con una alta velocidad de adquisición de las imágenes, es un área en crecimiento, debido a que la mayoría de los sistemas de Imagen Hiperespectral usados en laboratorio son caros, lentos e inapropiados por su configuración para realizar medidas en tiempo real.

De entre las limitaciones de los sistemas de Imagen Hiperespectral de bajo coste encontramos la calibración de los instrumentos y un bajo ratio señal/ruido (“signal-to-noise”, SNR). Pese a la ubicuidad de las aplicaciones de los sistemas de Imagen Hiperespectral y su desarrollo por laboratorios independientes y/o firmas comerciales, el desarrollo paralelo de la calibración, caracterización de las longitudes de onda y la caracterización espectral no se ha realizado al mismo ritmo. En contraste con los sistemas espectroscópicos convencionales, los sistemas de Imagen Hiperespectral capturan la información espectral en cada localización y, por tanto, no se puede manejar el hipercubo resultante bajo la asunción de que las repuestas de cada píxel individual son iguales. Esto se debe, principalmente, a las inconsistencias relacionadas con la transformación de la luz en valores digitales (A/D) y las diferencias de iluminación entre píxeles.

Pese a la abundante literatura relacionada con ambos sistemas de Imagen Hiperespectral (barrido lineal y barrido espectral), no hay trabajos que los comparen en idénticas condiciones. Antes se hará una descripción general de los componentes disponibles para los sistemas de imagen hiperespectral de barrido lineal, como el usado para la consecución de los objetivos.

4.4.2. Componentes de un sistema de imagen hiperespectral de barrido lineal.

Fuentes de luz.

Los dos principales tipos de fuentes de luz usados en los sistemas de barrido lineal son las fuentes de iluminación y las de excitación. Las luces con un amplio rango espectral son las mayoritariamente usadas. Su constitución espectral no puede cambiar durante las

interacciones luz-muestra. Las lámparas de cuarzo-tungsteno (QTH) producen un suave espectro en un amplio rango espectral desde el visible hasta el infrarrojo; son ampliamente usadas como fuentes de iluminación, tanto en transmitancia como reflectancia. Los diodos de emisión de luz (“light-emitting diodes” LEDs) también pueden ser usados.

Un haz de luz monocromática de alta intensidad con un amplio rango espectral puede excitar materiales biológicos durante las interacciones muestra-luz para generar luz de baja intensidad en un amplio rango espectral (p.ej., emisiones de fluorescencia y dispersión de Raman). Los láser son fuentes de excitación potentes ideales para las medidas Raman y de fluorescencia debido a su alta concentración de energía, perfecta direccionalidad y verdadera emisión monocromática. Los LEDs de banda estrecha son fuentes de excitación emergentes que ofrecen muchas ventajas, tales como una larga vida útil, bajo consumo de energía, baja generación de calor, pequeño tamaño, respuesta rápida, robustez e insensibilidad a la vibración. Otras fuentes de luz, como las lámparas de ultravioleta, las lámparas de arco y alta presión y las lámparas de vapores metálicos de baja presión pueden también usarse para excitar la muestra (Jianwei, 2017).

Los sistemas de barrido lineal usan generalmente un campo de visión (FOV, por sus siglas en inglés) instantáneo lineal. Tanto las lámparas lineales como las superficiales pueden proyectarse sobre las muestras para cubrir el FOV para evaluaciones de barrido lineal superficial.

Espectrógrafos.

Un espectrógrafo es un dispositivo óptico que separa espacialmente un haz de luz en sus distintas longitudes de onda. El espectrógrafo adquiere una línea de información espacial de un objetivo vía una ranura de entrada. Es el núcleo de sistema de Imagen Hiperespectral. Los difractores (gratings) de transmisión, por ejemplo, reúnen la luz gracias a una serie de lentes frontales y después la dispersan en un prisma de tipo PGP (prism-grating-prism), donde la dirección de propagación de la luz depende de la longitud de onda. La luz dispersada es proyectada en un detector a través de unas lentes traseras, creando una imagen 2-D: una dimensión es la espacial y la otra la espectral.

Hay espectrógrafos dispersivos para imagen disponibles comercialmente para distintas regiones espectrales, tales como ultravioleta y visible (UV-VIS: 250-500 nm), visible (VIS: 380-800 nm), visible e infrarrojo cercano (VNIR: 400-1000 nm), infrarrojo cercano (NIR: 900-1700 nm) e infrarrojo medio (MWIR: 3000-5000 nm). También hay disponibles espectrógrafos de imagen que trabajan en rangos espectrales más estrechos (p.ej., espectrógrafos Raman -770-980 nm) para aplicaciones más particulares con necesidades de mayores resoluciones espectrales (Jianwei, 2017).

Sensores de imagen.

Los sensores de imagen usados en sistemas de barrido lineal suelen colocarse con una dimensión paralela a la dirección de dispersión de las longitudes de onda para la obtención de la información espectral y la otra paralela a la ranura de entrada del espectrógrafo para la obtención de la información espacial. Los dispositivos de carga acoplada (charge-couple devices –CCD-) son actualmente los más ampliamente usados. Sus respuestas espectrales dependen directamente del material usado en su construcción. Los CCDs de silicón tienen una adecuada eficiencia cuántica (quantum efficiency –QE-) en infrarrojo cercano, visible e infrarrojo de onda corta (SWNIR) (400 - 1000 nm). Los CCD de indio, galio y arsenio (Indium gallium arsenide –InGaAs-) son los sensores usados mayoritariamente en la región NIR. Tienen una eficiencia cuántica (QE) plana y alta en la región de 900-1700 nm. Alterando la proporción de InAs y GaAs en el sensor, pueden usarse eficientemente en la región SWIR (1000-2500 nm). Para esta región hay también detectores como los de plomo-selenio (PbSe), indio-antimonio (InSb) y mercurio-cadmio-telurio (MCT) que funcionan eficientemente. Las aplicaciones hiperspectrales con bajas intensidades de luz necesitan CCDs de alto rendimiento, tales como los CCD multiplicadores de electrones (electron-multiplying CCDs, o EMCCDs) o CCDs intensificados (ICCEs). Fuera de los CCDs, las cámaras CMOS (semiconductores de óxido metálico complementarios) han emergido como sensores de imagen alternativos, haciendo mucho más rápida la transferencia de señal que los registros verticales y horizontales usados por los CCDs para desplazar las cargas eléctricas. Las cámaras CMOS son particularmente idóneas para las aplicaciones

hiperespectrales de alta velocidad, tal como las inspecciones para seguridad alimentaria o calidad en líneas de procesado (Jianwei, 2017).

4.2. Herramientas matemáticas para el análisis de imágenes hiperespectrales.

Una vez que se ha generado el hipercubo de una muestra, compuesto de un espectro para cada píxel, hay que tener en cuenta que estos espectros no representan a una sustancia, sino la suma de la contribución a la reflectancia/absorbancia de cada uno de los componentes que conforman la mezcla. Esta característica, junto a la elevada cantidad de información contenida en los espectros NIRS, hace necesario el uso de herramientas matemáticas para la extracción de información relevante de los datos. Estas herramientas las agruparemos en aquellas originarias del análisis NIRS clásico o quimiométricas, las originarias del análisis de imagen, más concretamente de patrones en las imágenes, y por último aquellas propias de la minería de datos.

En la figura 3 se indican los pasos y alternativas habituales en el análisis de una imagen hiperespectral. Se describe en primer lugar brevemente el proceso en su conjunto, para después detallar más cada una de las herramientas matemáticas. El primer paso consiste en la eliminación de los píxeles oscuros o muertos y píxeles saturados. Ambos son consecuencia de la ruptura de píxeles del CCD. Los primeros hacen referencia a píxeles que no captan señal alguna. Los segundos son píxeles mucho más sensibles que el resto y que, por tanto, se saturan antes. Los primeros se eliminan tomando una imagen de una referencia blanca, en general un *Spectralon* o cerámica blanca. Seguidamente, se localizan los píxeles con valores anormalmente bajos en dicha imagen. Éstos son sustituidos en la adquisición de cada imagen por el promedio de los cuatro píxeles colindantes para cada longitud de onda. Un proceso similar se sigue para los píxeles saturados, con la diferencia de que se localizan sobre una imagen sin señal, obtenida mediante el cierre del objetivo de la cámara.

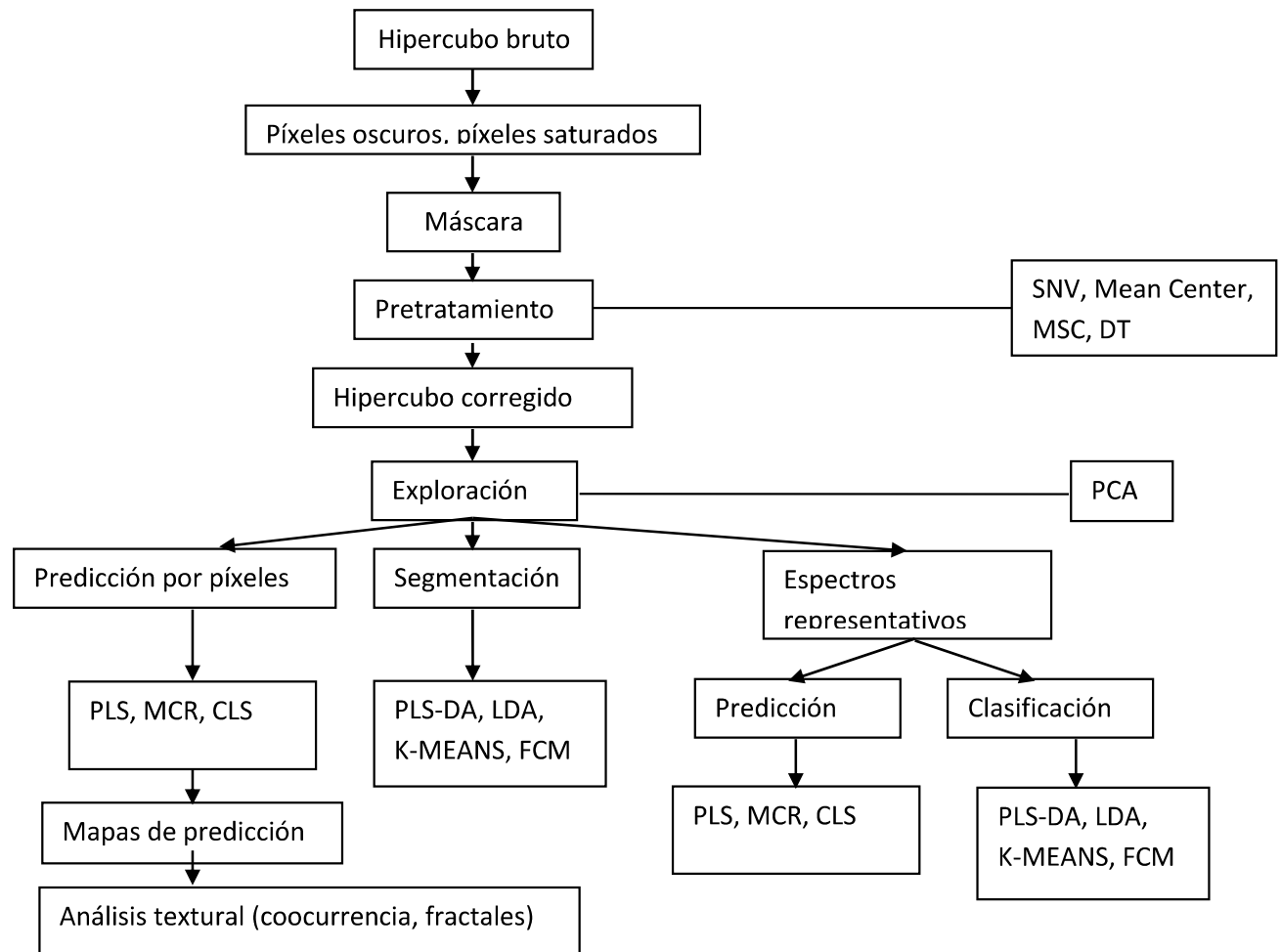


Figura 3. Esquema del análisis de una imagen hiperespectral.

Una vez corregidos los píxeles defectuosos de la imagen debido al deterioro de la cámara, y eliminado aquello que no interesa de la misma (el fondo de la imagen, por ejemplo) se corrigen aquellos defectos de la imagen asociados principalmente a efectos físicos, mediante una serie de tratamientos matemáticos que se denominan genéricamente pretratamientos de la señal. El resultado de estos pretratamientos es el hipercono corregido.

Antes de aplicar cualquier modelo predictivo o clasificador, se realiza un análisis exploratorio de las imágenes. Todos los algoritmos matemáticos usados tanto en análisis clásico de datos NIRS como de datos de Imagen Hiperespectral provienen del análisis multivariante, métodos estadísticos utilizados para determinar la contribución de varios factores en un simple evento o resultado. Se basan en la distinción entre múltiples variables independientes o variables explicativas, que pueden ser o no linealmente

independientes, y una o varias variables dependientes o variables respuesta. Estos métodos permiten la creación de espacios n-dimensionales diagonalizables.

La herramienta principal para este primer análisis es el Análisis de Componentes Principales (ACP o PCA). Con él, la gran cantidad de datos de cada hipercubo se organiza en torno a un sistema de ejes ortogonales a lo largo de las direcciones de máxima varianza de los datos. A partir de este análisis, se pueden determinar las relaciones entre distintas proyecciones de los datos sobre estos nuevos ejes y los componentes buscados en la muestra, permitiendo evaluar, por un lado, la bondad del pretratamiento realizado, y por otro, del mejor de los modelos a aplicar con posterioridad.

Tres son los enfoques que podemos adoptar en el análisis de un hipercubo: el predictivo, la segmentación y el de los espectros representativos. Se seguirá uno u otro dependiendo del objetivo del análisis. Así, por ejemplo, para la obtención de mapas de concentraciones se seguirá el enfoque predictivo; para la separación de la muestra en submuestras, la segmentación; y para el análisis según la quimiometría clásica, el enfoque de espectros representativos. Hay objetivos que pueden abordarse desde distintos enfoques, como puede ser la detección de contaminaciones. En efecto, la detección de un contaminante, es decir, una partícula con una señal espectral distinta al resto de partículas, podría determinarse bien por segmentación, quizás hasta el nivel de píxel, por clasificación de píxeles o por predicción, si se conoce la señal de una muestra de contaminante representativa. A continuación se expondrán los fundamentos de las principales herramientas aplicadas en el análisis de Imagen Hiperespectral. Se indicará en qué parte o partes de la presente tesis doctoral se usaron, así como los programas o paquetes estadísticos con los que se aplicaron.

4.2.1. Pretratamiento espectral.

Los pretratamientos espectrales o preprocesado son correcciones del espectro basadas en tratamientos matemáticos con la finalidad de aumentar la calidad de la señal instrumental reduciendo al máximo la información no asociada a la propiedad estudiada, como pueden ser desplazamientos de la línea base, ruido espectral, etc. (Kennard y

Stone, 1969). Se denomina ruido al componente espectral no relacionado con ninguna propiedad física o química de la muestra, como puede ser el ruido instrumental o de iluminación. Además, otras fuentes de variación como la temperatura y humedad pueden aportar interferencias a la señal analítica.

El Standard Normal Variate (SNV) corrige efectos de dispersión o scattering provocados por diferencias físicas entre muestras (Massart et al., 1997). El método se aplica individualmente a cada espectro y se obtiene un espectro de absorbancia media 0 y desviación estándar 1. El Detrend (DT) elimina la curvatura de la línea base causada por interacciones variables entre la humedad y efectos de partícula (Geladi et al., 1985). Suele usarse como complemento al SNV.

La corrección del efecto multiplicativo de dispersión (MSC) corrige la dispersión de la luz y los efectos espectrales aditivos y multiplicativos (Fearn et al., 2009). Pese a ser un pretratamiento equivalente al SNV, existen, no obstante, diferencias entre el SNV y MSC. El SNV tiene la capacidad de inducir estructuras curvadas en los diagramas de coordenadas (“score plot”) de los espectros de tratados, mientras que MSC tiene una tendencia a producir valores anómalos en los diagramas.

El Mean Center o centrado de la media extrae el valor medio de la variable para cada uno de los valores de la variable original. De esta manera, cada variable en la nueva matriz de datos (matriz centrada) presenta a una media igual a cero. El Mean Center hace una translación de las coordenadas originales de cero al valor medio de los datos. De esta manera, las desviaciones de la media son fáciles de visualizar, ya que esta parte de la información de las variables es considerada constante y puede ser eliminada.

En el presente trabajo de tesis doctoral, todos los pretratamientos citados fueron usados en todos los objetivos planteados. Todos ellos se aplicaron usando funciones programadas en Matlab R2008 o a través de su PLS toolbox R2008a, salvo en el objetivo 1, en el cual se usó el programa WinISI v. 1.5.

4.2.2. Exploración del hipercubo. Análisis por Componentes Principales y detección de anómalos.

El algoritmo de análisis paradigmático para la exploración tanto de señales espectrales individuales como señales espectrales en imágenes hiperespectrales es el Análisis de Componentes Principales (ACP o PCA). Se parte de la hipótesis de que existen una serie de fuentes de variación que influyen de manera importante en la variación del conjunto de datos observados. Estos pocos componentes explican la mayor parte de la información contenida en los datos, siendo redundante (correlación con otras variables) o irrelevante. El análisis PCA permite la transformación de este conjunto de datos en uno de variables no correlacionadas (sistema de referencia de ejes ortogonales).

Matemáticamente la matriz de datos original, D , es descompuesta en el producto de dos matrices de datos ortogonales, X y Y^T :

$$D=XY^T +E=D*+E$$

donde D es la matriz de datos experimentales, X la matriz de “scores”, Y^T la matriz de “loadings” y E la matriz de error (Fernández, 2003).

Se denomina “loading” a la proyección del antiguo sistema de referencia en el nuevo. Cuando se consideran datos espectrales los “loadings” pueden proporcionar información sobre la composición química de una muestra. Las proyecciones de las muestras en el nuevo espacio se denominan “scores”. Su valor exploratorio reside en que nos permiten la agrupación de las muestras de acuerdo con su similitud, así como la detección de muestras que presentan valores extremos. La matriz de error E contiene la parte de varianza no explicada por el modelo y que, en consecuencia, al menos hipotéticamente, es el resultado del error experimental.

La información más relevante del set de datos se concentra siempre en los primeros componentes principales generados (los que explican un mayor porcentaje de la varianza). Es por ello que los últimos componentes pueden omitirse sin una pérdida significativa de información. En general, el número de componentes seleccionado es aquel donde la adición de un nuevo componente no proporciona información relevante en el contexto del problema estudiado, generándose en caso contrario un sobreajuste del modelo a los datos.

El análisis exploratorio consiste en la extracción de significado de los componentes identificados y en la determinación del nivel aproximado de error contenido en los datos experimentales. Para ello, la visualización de los “scores” (X) en forma de imagen (cada “score” situado sobre cada píxel, como puede apreciarse en la figura 4 y de los “loadings” (Y^T) en forma de espectro (longitud de onda vs peso de cada uno de ellos) son extremadamente útiles.

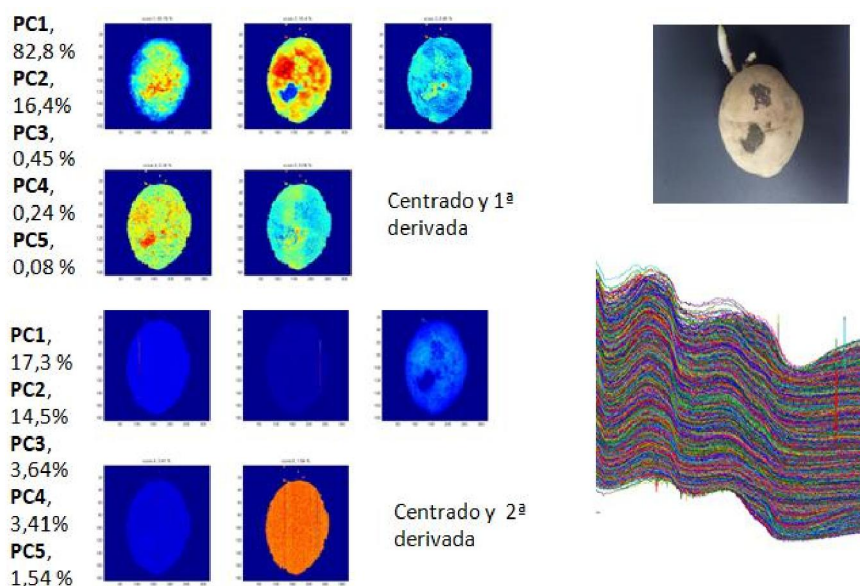


Figura 4. Imagen, espectros NIRS y “scores” tras Análisis de Componentes Principales de una patata contaminada con estiércol y tierra (elaboración propia).

La elección del número de componentes principales puede realizarse de varias maneras. Las dos usadas en la presente tesis consistieron en elegir el número de componentes principales que minimizaran el error típico de validación cruzada (SECV, Standard Error of Cross Validation) a la vez que se evita una desviación excesiva de error típico de calibración (SEC, Standard Error of Calibration) con respecto al SECV (en torno a un 20 %), para evitar sobreajustes.

Se usó PCA en la consecución de los cuatro objetivos de la presente tesis. En el primero de ellos se aplicó usando el programa WinISI 1.50. En los otros tres bien en Matlab R2008a directamente o a través de la PLS toolbox del mismo programa.

Otro uso del análisis PCA es el de la detección de espectros anómalos. Hotelling’s T^2 y Q-Residuals son estadísticos que ayudan a explicar cómo un modelo describe una muestra concreta. No sólo se puede usar con PCA. También puede usarse con otros modelos multivariantes (Partial Least Squares, Multivariate Curve Resolution, etc.).

Esta aplicación se estudió y aplicó con algo de mayor profundidad en la consecución del cuarto objetivo de la presente tesis [“*Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins*”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99].

El estadístico Hotelling’s T^2 fue propuesto originariamente por Hotelling (Hotelling, 1947). Mientras que el Hotelling’s T^2 representa la variación de cada muestra dentro del modelo, el Q-Residuals representa la magnitud de la variación que queda sin modelizar de cada muestra, o dicho de otro modo el error del modelo para cada muestra.

Hotelling’s T^2 se calcula como la suma de los cuadrados de los “scores” según la ecuación siguiente:

$$T_i^2 = x_i P_k \lambda^{-1} P_k^T x_i^T$$

donde x_i es el vector de respuestas para la muestra i , P_k es la matriz de “loadings” y λ es una matriz diagonal que contiene los valores propios asociados a los vectores propios incluidos en el modelo PCA. El límite de control del gráfico T^2 se calcula según:

$$T_{k,m,\alpha}^2 = \frac{k(m-1)}{m-k} F_{k,m-k,\alpha}$$

donde m es el número de análisis de la muestra que se han utilizado para construir el modelo PCA, k el número de PCs incluidos en el modelo PCA y $F_{k,m-k,\alpha}$ el valor de la distribución F para k y $m-k$ grados de libertad y una probabilidad α de cometer un error.

El gráfico T^2 monitoriza la distancia de una nueva medida al valor de referencia en el espacio reducido de los factores PCA. Permite detectar si la variación incluida en los PCs considerados es mayor que si solo influyera la variación aleatoria. Las muestras fuera de control poseen un valor de T^2 superior al límite, y aparecen más allá de la línea de control.

El estadístico Q-Residuals se calcula mediante la ecuación siguiente:

$$Q_i = e_i e_i^T = x_i (I - P_k P_k^T) x_i^T$$

donde e_i es la fila i de la matriz de los residuales (E), I es la matriz identidad, P_k la matriz de los k vectores de “loadings” incluidos en el modelo PCA y x_i es el vector fila de la respuesta de la muestra i . El límite de control del gráfico Q se calcula según:

$$Q_\alpha = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}}$$

donde: $\theta_i = \sum_{j=k+1}^n \lambda_j^i$ para $i = 1, 2, 3$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$ y k es el número de componentes principales retenidos en el modelo, λ_j el valor propio asociado al componente principal j , n el número total de componentes principales y c la STD normal correspondiente a un error (α) fijado. El valor de Q indica el grado de ajuste de cada muestra al modelo PCA y da una medida de la variación de la muestra no incluida en el modelo.

La representación de Hotelling $-T^2$ versus Q -residuals y ambos límites de control nos permite detectar anómalos. Éstos serían aquellas muestras con ambos valores por encima de sus respectivos límites, es decir aquellas muestras que estando lejos del centro muestral del modelo (Hotelling $-T^2$ por encima del límite de control) son mal explicadas por el modelo (Q -residuals por encima del límite de control). De todos modos, y como se tratará de analizar en el cuarto objetivo de la presente tesis, la eliminación de muestras espectralmente anómalas según este criterio sin conocer la causa de dicha anomalía comporta el riesgo de pérdida de información valiosa.

Otro modo de detectar la presencia de anómalos espectrales es mediante la distancia de Mahalanobis. Es utilizada para determinar la distancia de un punto en el espacio multivariante a un grupo de otros puntos. Valores de la Distancia Global de Mahalanobis (GH, Global Mahalanobis Distance), definida como la distancia entre una muestra y el centro de la población en el espacio definido por un análisis de componentes principales, superiores a 3 indican muestras espectralmente anómalas, al igual que muestras con valores de NH (Distancia de Mahalanobis de Vecinos o Neighbourhood Distance, definida como la distancia entre una muestra y aquella más cercana a ella en una PCA) superiores a 1-1,2, ya que son consideradas como muestras poco representadas en el colectivo de calibración (Shenk y Westerhaus, 1995).

4.2.3. Predicción.

Partial Least Squares (PLS).

Se conoce como regresión PCR o Regresión por Componentes Principales, la regresión lineal desarrollada entre la matriz de datos de los principales componentes obtenidos mediante PCA y la matriz de datos de los valores referencia o químicos. La principal debilidad de este algoritmo reside en que sólo usa la información espectral para calcular las nuevas variables, lo que no implica que sean las más apropiadas para realizar la predicción (Massart et al. 1997).

En cambio, la regresión PLS (Partial least Squares) o regresión por Mínimos Parciales calcula las nuevas variables ortogonales no sólo a partir de la matriz de datos espectrales, sino considerando la matriz de los valores de referencia. Podríamos decir que hace una rotación del sistema de referencia (Fig. 5) calculado mediante PCA para que sus ejes, direcciones de máxima varianza de los datos espectrales, coincidan con las direcciones de máxima varianza de los datos de referencia o químicos. Así, las nuevas variables (LV, Latent Variables) maximizan la varianza explicada de la matriz espectral (X), de forma que sea relevante para predecir los parámetros de la matriz con los datos de referencia (Y) (Martens y Næs, 1985).

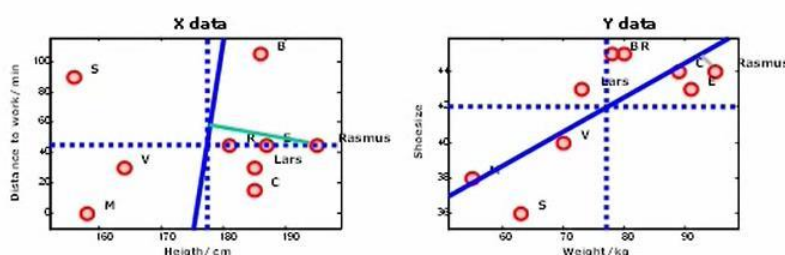


Figura 5. Rotación de los ejes de referencia tanto en la variable dependiente como en la independiente en el sentido de la máxima varianza (Fuente: models.life.ku.dk).

Existen dos variantes del algoritmo: PLS1 y PLS2 (Manne, 1987). Mientras que PLS2 tiene en cuenta para el cálculo de las LV varios parámetros de referencia a la vez, PLS1 se basa en el desarrollo de un modelo PLS para cada uno de los parámetros independientemente.

El método de elección del número de variables latentes en PLS fue el mismo que el expuesto previamente para la elección del número de componentes principales en PCA basado en el SECV y el SEC.

El PLS fue aplicado, usando la PLS toolbox en Matlab R2008a, en el segundo y el tercer objetivo de esta tesis doctoral [*“Differentiation of beef and fish meals in animal feeds using chemometric analytic models”*, *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)] y [*“Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”*, *Journal of Food Engineering* 153 (2015) 117-123)].

Dos de los principales problemas a los que se enfrentan la mayoría de los intentos de predicción química a nivel de píxel en Imagen Hiperespectral son la pérdida de gran parte de la información disponible por el uso de espectros medios, y la artificialidad y sobreajuste si se usan la enorme de espectros disponibles. Existe, no obstante, una validación específica para modelos hiperespectrales puesta a punto por Burger y Geladi (2006) que se usó en la presente tesis, concretamente en su segundo objetivo [*“Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”*, *Journal of Food Engineering* 153 (2015) 117-123)].

La construcción de modelos de predicción con Imagen Hiperespectral implica la regression entre hipercubos y datos externos a la misma. Debido a que los datos externos o de referencia no están disponibles para cada píxel de la imagen, ha de considerarse un único valor global que los represente a todos. Este valor global ha de estar relacionado con la señal espectral representativa del hipercubo, por ejemplo el espectro medio. Burger y Geladi (2006) recomiendan la aplicación de la validación cruzada tanto a los datos de calibración como a los de predicción para minimizar el exagerado desequilibrio entre las decenas de datos de predicción frente a los millones de millones de datos de calibración. Para ello proponen los siguientes estadísticos, pensados específicamente para Imagen Hiperespectral: el RMSEP global (RMSEPG) y la Desviación Estándar Global (Global Standard Deviation. SG):

$$RMSEPG = \left[\frac{r^t r}{P} \right]^{1/2}$$

$$SG = \left[\frac{\sum (s^2 (L - 1))}{\sum (L - 1)} \right]^{1/2}$$

donde, r es el vector de desviación de cada imagen; P es el número total de hipercubos; s es la desviación estándar de cada imagen; L es el número de píxeles por cada imagen.

SG representa el nivel de varianza esperada dentro de una sola imagen, mientras que RMSEPG representa la precisión esperada en la predicción de una nueva imagen. Para seleccionar el mejor modelo se representa gráficamente RMSEPG vs SG, representando cada punto del gráfico un modelo distinto (pretratamientos y número de variables latentes distintos). Para un modelo mal ajustado, el RMSEPG será algo mayor. A medida que nuevas variables latentes son añadidas, el ajuste global mejora y el RMSEPG se reduce. En cambio, cualquier varianza adicional incluida no tiene por qué correlacionarse con los valores de referencia, lo que contribuirá a un incremento en la varianza de los valores predichos, SG.

Multivariate Curve Resolution (MCR) y Classical Least Squares (CLS).

La base de los métodos MCR (*Multivariate Curve Resolution*) y el CLS (*Classical Least Squares*) es la misma. Ambos asumen el comportamiento lineal de la absorbancia con respecto a la concentración (Ley de Beer–Lambert), y que la suma de las absorbancias individuales para cada componente es igual a la absorbancia total para cada píxel (Eq. (1)).

$$a_{xy} = c_{1xy}\epsilon_1 + c_{2xy}\epsilon_2 + c_{3xy}\epsilon_3 + \dots + c_{fxy}\epsilon_f \quad (1)$$

donde a_{xy} es la absorbancia de cada píxel xy a cada longitud de onda (λ); c_{fxy} y ϵ_f son la concentración y la absorptividad molar, respectivamente, para cada componente f de la superficie.

Matemáticamente hablando, ambos son modelos bilineales, de modo que es obligatorio un paso previo de abatimiento de la estructura tridimensional en otra bidimensional. Consecuentemente, las dimensiones de la imagen sin abatir son $(XY \times \lambda)$.

La principal diferencia entre el CLS y el MCR-ALS (Alternating Least Squares) reside en el modo de obtención de la información cuantitativa. El CLS está orientado a obtener la concentración de cada componente por regresión directa de D , usando los espectros puros (Amigo y Ravn, 2009). Pese a la facilidad de calcular esta matriz de concentraciones, el CLS necesita los espectros puros de cada componente y la ausencia de otras fuentes de variación en la muestra. Por ejemplo, la interacción entre componentes, la distribución de la humedad, etc. (De Juan et al., 2004).

$$D=CS^T+E \quad (2)$$

donde E ($XY \times \lambda$) contiene el error experimental.

La principal diferencia entre el CLS y el MCR-ALS es que el segundo trabaja iterativamente optimizando las matrices, pudiendo absorber fuente menores de sensibilidad. MCR-ALS necesita una estimación inicial así como limitaciones basadas en el conocimiento químico sobre la muestra. Usar el espectro puro de los componentes, si está disponible, como estimación inicial se convierte en una de los modos más atractivos de incluir información sobre la muestra.

Alguna de las restricciones químicas que se pueden usar son: (1) la no negatividad de la concentración y/o los espectros y (2) cada píxel ha de cumplir con el balance constante de masa de 1. La principal desventaja del MCR-ALS en la imagen química es la ausencia de selectividad en la superficie (Amigo et al., 2008). Esta falta de selectividad puede ser asociada al problema de la deficiencia de rango. MCR-ALS solo trabaja correctamente si el rango, es decir el número de componentes matemáticas, es igual al número de componentes químicas de la muestra. Para superar este inconveniente, se ha propuesto el aumento de la imagen original con más imágenes. Este método, propuesto por de Juan et al. (2004), trata de analizar bastantes muestras simultáneamente, beneficiándose unas de otras. La muestra (D_2) es aumentada con una imagen sin problemas de limitación de rango (rango completo o imagen estándar, D_1) en forma de columna (Eq. (3)):

$$\begin{matrix} D_1 \\ D_2 \end{matrix} = \begin{matrix} C_1 \\ C_2 \end{matrix} S^T + \begin{matrix} E_1 \\ E_2 \end{matrix} \quad (3)$$

donde D_1 ($X_1Y_1 \times \lambda$) representa la imagen de rango completo y D_2 ($X_2Y_2 \times \lambda$) es la imagen de rango deficiente (la imagen problema). X_1 , Y_1 y X_2 , Y_2 representan las

dimensiones de la primera y segunda imagen desdoblada respectivamente. C_1 ($X_1Y_1 \times F$) y C_2 ($X_2Y_2 \times F$) son las superficies de concentraciones obtenidas simultáneamente para cada imagen y ST ($F \times \lambda$) es la traspuesta de los espectros globales E_1 ($X_1Y_1 \times \lambda$) y E_2 ($X_2Y_2 \times \lambda$) son las matrices de residuales.

Esta imagen estándar se compone de secciones específicas de cada componente puro y puede crearse fácilmente a partir de imágenes de los componentes puros. Trabajando con este tipo de muestra artificial, la información modelada por el MCR-ALS es mucho más valiosa y fiable a la vez, ya que la variabilidad experimental también se refleja en la imagen aumentada, permitiendo al MCR-ALS obtener información selectiva en cada iteración y así minimizar la ausencia de selectividad en la muestra objetivo.

El MCR y el CLS se probaron en la consecución del tercer y cuarto objetivo de la presente tesis doctoral sin obtener resultados satisfactorios.

4.2.4. Segmentación y/o clasificación.

Partial Least Squares Discriminant Analysis (PLS-DA)

Durante el desarrollo de un modelo discriminante PLS, se conocen *a priori* los grupos a los cuales pertenecen las muestras del colectivo de calibración (pueden también generarse grupos artificiales a partir de datos continuos).

Los pasos que sigue el algoritmo son los mismos que se han descrito anteriormente para el modelo PLS, con la diferencia de que la matriz de valores de referencia es una matriz artificial con las muestras de todas las clases y unas nuevas variables, tantas como clases existan. Cada variable indica la pertenencia o no de una muestra a un determinado grupo mediante unos o ceros, respectivamente. Una vez obtenido el modelo se predice el valor de cada variable ficticia para cada muestra. El porcentaje de muestras correctamente clasificadas en las clases a las que pertenecen indica la bondad del modelo desarrollado (Goel et al., 2003).

Este algoritmo fue usado en la consecución del primer y cuarto objetivos de la presente tesis [*“Grading and color evolution of apples using RGB and hyperspectral imaging*

vision cameras”, Journal of Food Engineering 113 (2012) 281-288)] y [“Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, Chemometrics and Intelligent Laboratory Systems, 172 (2018) 90-99].

El algoritmo permite generar varias variables discriminatorias correspondientes a las clases existentes en el colectivo de calibración (Riccioli, 2011). El algoritmo incorpora el factor de incertidumbre que está relacionado con el estadístico *T de Student* para un número de grados de libertad igual a infinito. Parte de la realización de un modelo PLS para reducir la dimensión de la matriz y establecer las fronteras entre las clases. Posteriormente, hace un análisis discriminante y clasifica las nuevas muestras según el modelo creado. Para obtener un código de identidad de las clases usa como matriz *Y* (variables dependientes), una matriz de variables ficticias. A partir de esta matriz de variables ficticias se crea un modelo PLS con variables binarias para cada clase como respuesta. Su función es identificar variables latentes características que puedan ser utilizadas para integrar las muestras a una clase.

Asociado a este algoritmo, se usó otro parámetro en el cuarto objetivo de la presente tesis para conocer las variables más influyentes: la Proyección de Variables Importantes (VIP, del inglés Variables Important Projection) que identifica las longitudes de onda (λ) claves en el análisis (Jiang et al., 2007).

Linear Discriminant Analysis (LDA)

El algoritmo LDA es muy cercano al análisis de la varianza (ANOVA) y al análisis de regresión en su intento de expresar una variable dependiente como la combinación lineal de otras medidas. No obstante, mientras que el ANOVA usa variables categóricas independientes y variables continuas dependientes, éste usa variables continuas independientes y variables categóricas dependientes. Una observación es clasificada en un grupo si la distancia de Mahalanobis a su centro es la mínima, asumiendo que las matrices de covarianza son las mismas para todos los grupos. Este algoritmo fue usado en la consecución del primer objetivo [*“Grading and color evolution of apples using*

RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288)], mediante el paquete estadístico Minitab.

K-MEANS y Fuzzy Cluster Means Analysis (FCM).

El análisis de clusters realiza el análisis de las imágenes hiperespectrales segmentándolas en función de la información química de sus píxeles de manera no supervisada, es decir, sin ningún set de entrenamiento previo. Destacan los dos algoritmos de segmentación por cluster no jerárquicos K-Means y FCM (Amigo et al., 2008).

El K-Means asigna cada píxel de la imagen al clúster cuyo centro está más cercano, minimizando la suma de las distancias cuadradas de cada píxel a sus correspondiente clúster. Se trata de un proceso iterativo que va convergiendo hacia la situación óptima. La principal ventaja de este algoritmo es la simplicidad que le permite manejar grandes conjuntos de datos. La principal desventaja es que hay riesgo de convergencia en un mínimo local.

$$J = \sum_{K=1}^K \sum_{mn=1}^{MN} \|x_{mn}^{(k)} - m_k\|^2$$

donde $x_{mn}^{(k)}$ es el mn-ésimo píxel y m_k es el centro del cluster k.

Los pasos que sigue el algoritmo en su funcionamiento es el siguiente:

- 1) Elige el número de clusters k.
- 2) Genera los k clusters y determina sus centros.
- 3) Asigna a cada píxel el centro más cercano.
- 4) Calcula J y recalcula nuevos centros de clusters.
- 5) Repite 3 y cuatro hasta la convergencia del criterio.

En el Fuzzy C-means a cada píxel se le asigna un grado de pertenencia a todos los clusters simultáneamente. FCM permite a los píxeles de borde pertenecer a un clúster con un menor nivel de pertenencia que los píxeles del centro. La función de pertenencia

asigna a cada píxel un valor entre 0 y 1. La suma de todos los coeficientes de todos los clúster para cada píxel debe ser la unidad.

El coeficiente de pertenencia u_{mnk}^g es calculado para cada x_{mn} píxel de tal modo que cada coeficiente esté comprendido entre 0 y 1, y la suma de todos los coeficientes de un píxel sea 1.

$$u_{mnk}^g = \left[\sum_{i=1}^c \left(\frac{\|x_{mn}^{(k)} - m_k\|^2}{\|x_{mn}^{(k)} - m_i\|^2} \right)^{1/(g-1)} \right]^{-1}$$

$$\forall x_{mn} \rightarrow \sum_{k=1}^K u_{mnk}^g(x_{mnk}) = 1$$

donde g es la constante conocida por “fuzzifier” o difusora, que determina la difusividad del resultado de agrupación. Un valor recomendado de g es 2, que indica que los coeficientes están linealmente normalizados para que su suma sea 1.

Ahora, la función J es calculada del siguiente modo:

$$J = \sum_{K=1}^K \sum_{mn=1}^{MN} u_{mnk}^g \|x_{mn}^{(k)} - m_k\|^2$$

Y el centro del cluster m_k es calculado como la media ponderada:

$$m_k = \frac{bf \sum_{mn=1}^{MN} u_{mnk}^g x_{mn}}{\sum_{mn=1}^{MN} u_{mnk}^g}$$

El proceso iterativo es similar al que sigue el K-Means. La principal diferencia, además del grado de pertenencia, es que FCM minimiza también la varianza intra-clúster. También en este algoritmo se corre el riesgo de alcanzar un mínimo local. Este segundo algoritmo (Fig. 20) se usó en la comunicación [“*Methodology for marbling characterisation of pure Iberian cured loin based on hyperspectral imaging*”. Comunicación oral 16th International Conference on Near Infrared Spectroscopy. 2013.”](Figura 6), paso previo para la consecución del segundo objetivo [La consecución del objetivo queda reflejada en la publicación: “*Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging*”, *Journal of Food Engineering* 153 (2015) 117-123)].

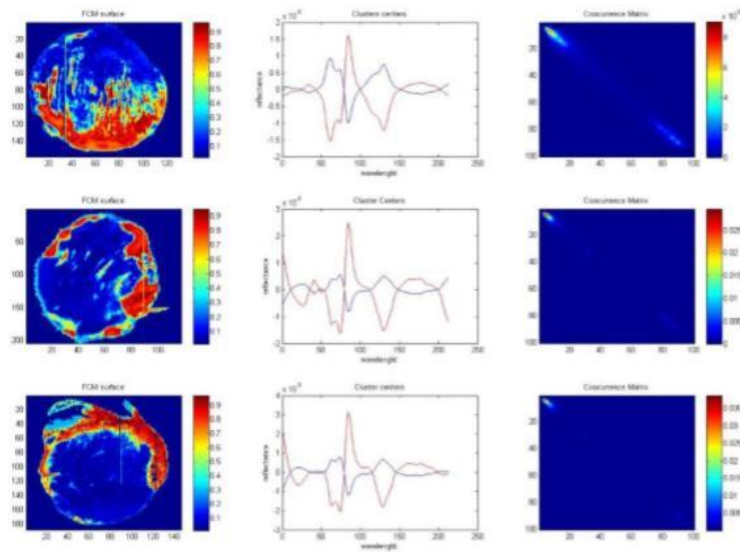


Figura 6. Imagen de la clasificación realizada por el FCM, espectros centrales de las clases encontradas y matrices de coocurrencia de dichas imágenes en rodajas de caña de lomo.

4.2.5. *Análisis textural*

Matriz de coocurrencia

Una matriz de coocurrencia (GLCM), como las mostradas en la Fig. 6, es una matriz cuadrada cuyos elementos corresponden a la frecuencia relativa de ocurrencia de pares de valores de gris de píxels separados por una cierta distancia en una dirección determinada (0° , 45° , 90° or 135°) (Haralick, 1973). El procedimiento estándar extrae los atributos texturales de imágenes monocromáticas. Cuando es aplicada a una imagen hiperspectral, la extracción suele llevarse a cabo usando imágenes monocromáticas en una banda específica o en proyecciones sobre los componentes principales. Existen autores, como Huang et al. (2014), que han propuesto métodos para aplicar el análisis textural a imágenes multiespectrales e hiperspectrales, usando metodologías alternativas al análisis de componentes principales.

La información de la matriz de coocurrencia se sintetiza en los llamados parámetros texturales, como el contraste (C), la homogeneidad (H), la energía (E) o segundo momento angular y la correlación (CR):

$$C = \sum_{i,j} |i - j|^2 p(i, j)$$

$$H = \sum_{i,j} \frac{p(i, j)}{1 + (i - j)^2}$$

$$E = \sum_{i,j} p(i, j)^2$$

$$CR = \sum_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] p(i, j)$$

donde i y j son reflectancias de píxeles y $p_{i,j}$ la probabilidad de coocurrencia de dichas reflectancias.

Esta metodología se usó en la consecución del segundo objetivo “*Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging*”, *Journal of Food Engineering* 153 (2015) 117-123] y el cuarto [La consecución del objetivo queda reflejada en la publicación: “*Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins*”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99].

Fractales

Las estructuras irregulares de muchos alimentos eluden la descripción mediante la tecnología convencional. El análisis de fractales es una poderosa herramienta para caracterizar tales geometrías y patrones irregulares en términos cuantitativos. La dimensión fractal puede ser utilizada para relacionar las características morfológicas de un alimento o muestra con propiedades funcionales e incluso sensoriales, facilitando el desarrollo, el control del proceso y el aseguramiento de la calidad en la industria alimentaria.

El concepto de fractal fue propuesto por Mandelbrot en 1983 para describir dimensiones “entre” las convencionales 1, 2 y 3 y estructuras que no son ni líneas euclidianas, ni superficies ni sólidos. La dimensión fractal indica el grado en que una imagen o el perfil de un objeto se desvían de la regularidad. Por ejemplo, una dimensión fractal de 2 a 3 describe la capacidad de llenado de volumen de una superficie muy rugosa. En los objetos euclidianos — esferas o cubos, por ejemplo — existen proporcionalidades constantes entre la dimensión lineal (el radio o el lado, por ejemplo), el área y el volumen, dadas por:

$$\text{Longitud} \propto \text{Área}^{1/2} \propto \text{Volumen}^{1/3}$$

Sin embargo, para estructuras e imágenes altamente convolucionadas, estas relaciones están gobernadas por diferentes exponentes. Mandelbrot (1980) cita como ejemplos el cerebro humano, cuya superficie está relacionada con el volumen^{D/3} con $D > 2$ o la forma fragmentada de muchas islas, cuya longitud está relacionada con el área^{D/2} con $D > 1$.

Una característica de los fractales matemáticos es su auto-similaridad, el atributo de tener la misma apariencia a todas las escalas. En los materiales reales o fractales naturales esta propiedad se da en un rango de escalas determinadas por restricciones físicas o biológicas. Un cierto grado de auto-similaridad se da en los helechos, en la coliflor o el brócoli, por ejemplo. Estos objetos son auto-similares en un cierto rango de escalas; no lo son a nivel celular. Mucho menos a nivel molecular. No obstante, los objetos naturales pueden ser descritos en términos de su dimensión fractal, que puede servir como un índice de rugosidad o tortuosidad morfológica.

Otra característica es la dificultad de calcular determinadas dimensiones. El área exacta o el perímetro de objetos o imágenes altamente convolucionadas dependerá de la escala tomada. Richardson's demostró que la longitud de costa se incrementa progresivamente a medida que la medida de referencia disminuye (Goodchild, 1980).

En la presente tesis, como alternativa al análisis por coocurrencia, se probó la caracterización de imágenes mediante el cálculo de su dimensión fractal usando el programa FracLab 2.2 tanto en el objetivo 4 [*“Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”*, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99], como en el 2 [*“Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”*, *Journal of Food Engineering* 153 (2015) 117-123]. Pese a no obtenerse éxito en dichas aplicaciones (las pruebas no se encuentran reflejadas en dichas publicaciones), el potencial de esta técnica en combinación con la imagen hiperespectral merece desarrollos y pruebas futuras fuera del ámbito de la presente tesis.

4.2.6. Árboles de regresión.

Cualquier intento de desarrollar la inteligencia artificial ha de incluir la comprensión del aprendizaje. El aprendizaje automático provee una metodología potencial para la construcción de sistemas de alto rendimiento. La investigación sobre aprendizaje se ha realizado en el ámbito del desarrollo de sistemas auto-mejorables, el conocimiento estructurado en forma de conceptos, las redes discriminantes y el desarrollo de reglas de decisión. Dentro de éstos últimos se encuentran los llamados árboles de clasificación y regresión.

Los árboles de clasificación y regresión son algoritmos de aprendizaje automático para la predicción a partir de gran cantidad de datos (Quinlan, 1986). Son obtenidos por división repetitiva del conjunto de datos de partida, que se ajusta a un modelo simple de predicción basado en la toma de decisiones sencillas en cada partición. Como resultado, la cadena de divisiones o decisiones sucesivas puede representarse gráficamente como

un árbol de decisión. Las reglas de poda del árbol cambian de un subtipo de algoritmo a otro. Las reglas usadas en la consecución del cuarto objetivo de la presente tesis [*“Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”*, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99] fueron las contenidas en el paquete “rpart”, del programa Rattle (entorno R). Éstas fueron: número mínimo de muestras por nodo, mínimo número de muestras resultado de una división y la complejidad, que expresa el número máximo de desarrollo permitido del árbol.

4.3. Aplicaciones.

Existen trabajos de desarrollo y optimización de equipos de Imagen Hiperespectral en múltiples y variados productos (frutas y hortalizas, legumbres y cereales, carnes, productos lácteos, ovoproductos, harinas, etc). Se realizará a continuación una breve revisión bibliográfica de las aplicaciones realizadas con los productos objeto de estudio en la presente tesis: frutas y hortalizas (objetivo 1); productos de origen animal - carnes, con especial mención a las carnes procesadas (objetivo 2); harinas proteicas de origen animal (objetivos 3 y 4)-.

4.3.1. Frutas y hortalizas.

En la cadena alimentaria, las frutas y hortalizas, hasta su llegada al consumidor tienen el riesgo de sufrir daños mecánicos, térmicos, por contaminación y por patógenos. Muchos de ellos no son fáciles de reconocer a simple vista. El fundamento de la detección suele ser la variación en la composición química del producto — los contenidos de humedad y sólidos solubles, la concentración de azúcar, la acidez — o bien la predicción de sus características físicas, con la variación de la dispersión de la luz cuando ocurren daños mecánicos, así como la aparición de sustancias ajenas a las que componen el producto en el caso de contaminaciones (Siche et al., 2015). En otras ocasiones, la determinación de la composición química o ciertas características físicas del producto son interesantes con el fin de mejorar su normalización y/o clasificación. Resulta interesante clasificar los frutos por ° Brix, color, acidez o firmeza para controlar la calidad de su procesado o manipulado, y su presentación por categorías más nítidamente definidas al consumidor. Por último, y más próximo a la especificidad de las cámaras hiperespectrales, puede resultar interesante, también con fines de controlar mejor el proceso de producción, bien en campo, bien en planta, la determinación de la distribución de estas propiedades en cada pieza; sobre todo cuando éstas van a ser troceadas, tal y como ocurre, por ejemplo, en las plantas de procesado de frutas y verduras para conservas, confituras, mermeladas o cuarta gama.

Se han realizado con éxito ensayos con Imagen Hiperespectral para la detección temprana de defectos mecánicos en mango (Vélez et al., 2014), fresas (Nagata et al., 2006), manzanas (ElMasry et al., 2008; Lu R., 2003; Xing et al., 2005), pepinos (Ariana et al., 2006) y tomates (Lee et al., 2014), entre otros. La detección de la contaminación fecal se ha realizado con éxito usando Espectroscopía de Fluorescencia en manzana (Kim et al., 2007). Trabajos más específicos que el anterior afinan, detectando la contaminación por *Escherichia coli*, por ejemplo en espinacas envasadas (Siripatrawan et al., 2011) o los daños producidos por *Burkholderia cepacia* en cebollas.

Otro tipo de daños sobre los que se ha trabajado con Imagen Hiperespectral son los daños por frío. Se han puesto a punto metodologías para detectarlos en manzanas (ElMasry et al., 2009), pepinos, (Cheng et al., 2004; Liu et al., 2005) y champiñones (Gowen et al., 2009).

La absorción de la luz, que es mayor cuanto mayor es el contenido en clorofila, permite también la clasificación en función del grado de madurez en frutos como plátano (Rajkumar et al., 2012), manzana (ElMasry et al., 2008), tomate (Qin J y Lu R., 2008) o melocotón (Lu y Peng, 2006). Durante el proceso de maduración, la degradación de la clorofila superficial tiene lugar a la vez que desaparecen otros pigmentos coloreados (carotenoides, amarillos y antocianos, rojos). Dos colores aparecerán, pues: el amarillo de fondo y el rojo en superficie. El color de la manzana *Golden Delicious*, objeto de estudio en el primer objetivo de la presente tesis, varía del verde al verde amarillento combinado con un rosado rubor. Con el fin de aprovechar la mayor cantidad de información aportada por las cámaras hiperespectrales en comparación con las cámaras RGB, tanto desde el punto de vista de la evolución del color como desde el punto de vista de la variación química de la composición de la manzana durante su maduración, el primer objetivo de la presente tesis fue evaluar el potencial de un equipo de imagen hiperespectral para la categorización de manzanas *Golden Delicious* [“Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288]

Otro aspecto a considerar en cuanto a la calidad de frutas y hortalizas es la firmeza, relacionada también con la luz dispersada o efecto “scatter”. Se ha conseguido medir con acierto en tomates (Qin y Lu, 2008), plátanos, manzanas (Noh y Lu, 2007; Peng y Lu, 2008), melocotones (Lu y Peng, 2006) y fresas (Rajkumar et al., 2012).

Algunos parámetros químicos, relacionados con la calidad, como el pH o la concentración de sólidos solubles, han sido evaluados con éxito en fresa (ElMasry et al., 2007), manzana (Menesatti et al., 2009; y Zhao et al., 2009) o naranja (Liu et al., 2008). Otros compuestos como los fenólicos han sido determinados en vino (Morata et al., 2015) y uva (Nogales-Bueno et al., 2014; Nogales-Bueno et al., 2015).

En la Tabla 1, adaptada y completada a partir de Siche et al. (2015) se hace un resumen de dichas aplicaciones y algunas más, indicando el producto, los parámetros medidos, el rango espectral, el coeficiente de determinación y el porcentaje de muestras correctamente clasificadas por los modelos generados.

Tabla 1. Aplicaciones de equipos hiperspectrales en frutas y hortalizas. Se indica el producto, el objeto del análisis, el rango espectral, la precisión de la predicción o clasificación y la referencia.

<i>Producto</i>	<i>Parámetro</i>	<i>λ (nm)</i>	<i>Precisión (R^2 o %)</i>	<i>Referencia</i>
Arándano	Detección de cuerpos extraños	400-720	100 %	Tsuta, et al., 2006.
Arándano	Detección de hojas y tallos	1000-1400	100 %	Sugiyama et al., 2010.
Cereza amarga	Infestación por insectos	590-1550	96 %/85 %	Xing, et al., 2008.
Uva	Fenoles totales	950-1650	0,89/0,8	Nogales-Bueno et al., 2014.
	°Brix	950-1650	0,99/0,95	Nogales-Bueno et al., 2014.
	Acidez tritratable	950-1650	0,98/0,93	Nogales-Bueno et al., 2014
	pH	950-1650	0,94/0,94	Nogales-Bueno et al., 2014
	Contenido total de fenoles extractables	900-1700	0,82	Nogales-Bueno et al., 2015.
	Contenido de antocianos	900-1700	0,94	Chen et al., 2015.

	Discriminación entre distintas variedades	400-1000	86 %	Nogales-Bueno et al., 2015
Fresa	Daños mecánicos	950-1650	90,7 %	Nagata et al., 2006.
	Humedad, sólidos solubles y pH	650-1000	0,87, 0,80 y 0,92	ElMasry et al., 2007.
Mango	Daños mecánicos	650-1100	91.4 %	Vélez et al., 2014.
Manzana	Daños mecánicos	400-1000	94 %	ElMasry et al., 2008.
	Defectos	954-1350	Discriminación defectos en la piel	Nicolaï, et al., 2006.
	Detección de contaminación por heces	675-950	100 %	Liu, et al., 2007.
	Firmeza y sólidos solubles	500-1000	0,65 y 0,71	Lu, 2007.
	Firmeza y sólidos solubles	450-1000	0,89 y 0,88	Peng, 2008.
	Harinosidad	600-1000	93 %	Huang y Lu, 2010.
	Daños mecánicos	400-1000	86 %	Xing et al., 2005.
	Daños mecánicos	900-1700	94 %	Lu, 2003.
	Firmeza	500-1040	0,74	Noh y Lu, 2007.
	Firmeza/sólidos solubles	450-1000	0,89/0,88	Peng y Lu R, 2008.
	Firmeza/sólidos solubles	500-1000	0,84/0,86	Qin et al., 2009.
	Daños por frío	400-1000	98.4 %	ElMasry et al., 2009.
	Contenido en	685-	0,91	Zhao et al., 2009.

	azúcares		900		
	Índice de almidón		1000-1700	80,8 %	Menesatti et al., 2009.
	Contaminación fecal		400-1000	100 %	Kim et al., 2007.
Pera	Contenido en agua y azúcar		450-1000	99 % y 94 %	Hong et al., 2007.
Pomelo	Antracnosis		450-930	93.3 y 96.7 %	Zhao et al., 2010.
Melocotón	Firmeza		5000-1000	0,58 y 0,77	Lu y Peng, 2006.
Naranja	Sólidos solubles		700-1100	0,96 y 0,998	Liu et al., 2008
Plátano	Sólidos solubles		400-1000	0,85	Rajkumar et al., 2012.
	Humedad		400-1000	0,87	Rajkumar et al., 2012.
	Firmeza		400-1000	0,91	Rajkumar et al., 2012.
Pepinos	Daños mecánicos		900-1700	82/93 %	Ariana et al., 2006.
	Daños internos		400-1000	99 %	Ariana et al., 2008.
	Daños por frío		447-951	93 %	Cheng et al., 2004.
	Defectos		900-1700	82-93 %	Liu et al., 2006.
	Daños por frío		447-951	90 %	Liu et al., 2005.
Tomate	Firmeza y madurez		500-1000	0,66	Qin y Lu R 2008.
	Daños mecánicos		1000-1700	96.4 %	Lee et al., 2014
	Deteccion de infestantes		385-810	95 %	Staab, et al., 2009.
	Clorofila	en	400-	0,839	Ding, et al., 2011.

	hoja	800			
Champiñones	Daños por frío	400-1000	0,98		Gowen et al., 2009.
	Defectos	400-1000	90 %		Gowen et al., 2008
Cítricos pre-cosecha	Estado de madurez	269-1042	80 % y 89%		Okamoto et al., 2009.
Mandarina	Putrefacción	460-1020	91v%		Gómez-Sanchis et al., 2008.
Dátil	Tamaño y defectos en la piel	750-1200	87 %		Lee, et al., 2008.
Cebollas	<i>Burkholderia cepacia</i>	950-1650	0,87		Wang et al., 2012
Espinacas	<i>Escherichia coli</i>	400-1000	0,97		Siripatrawan et al., 2011
Patatas	Predicción del tiempo de cocción	400-1000	0,94/0,96		Nguyen et al., 2011
Col	Contaminación microbiana	700-1100	0,95		Suthiluk et al., 2008

4.3.2. Productos de origen animal.

La industria cárnica está implementando tecnologías que permiten la evaluación y el control de la calidad más rápidamente, con menos costes y minimizando el error humano. Las propiedades de la carne más frecuentemente estudiadas en el ámbito de la visión hiperespectral son la terneza, la jugosidad, el veteado, la contaminación por microbios y la composición química. La tecnología hiperespectral se ha mostrado como una técnica de gran recorrido debido a la relación existente entre la estructura muscular y las propiedades bioquímicas del músculo, que incluyen, entre otras, el pH (que influye en la actividad de las enzimas proteolíticas), la longitud del sarcómero (que indica el grado de contracción del músculo), el grado de proteólisis y la cantidad e insolubilidad del tejido conectivo y su composición (Siche et al., 2015).

En este sentido, se han realizado estudios para la predicción de la terneza en carne, correlacionando la señal espectral de imágenes hiperespectrales con la medida de terneza obtenida por instrumentos clásicos (ElMasry et al., 2012). Estas metodologías se han establecido tanto para carne fresca como procesada. Se han usado distintos rangos espectrales (400-1000 y 900-1700 nm) para la predicción de la terneza de carne cocida después de 14 días de maduración. Nagathan et al. (2008) categorizaron muestras de carne según tuvieran una terneza suave, media o dura, con porcentajes de muestras correctamente clasificadas del 96,4 %. Otros parámetros como el color, la frescura, la pérdida de humedad o el pH han sido abordados en otros trabajos (Barbin et al., 2013), con coeficientes de determinación que oscilan entre 0,77 y 0,98 según el parámetro.

La visión hiperespectral es una tecnología idónea para la evaluación de la distribución de distintos atributos químicos en los productos cárnicos. La grasa y la humedad varían entre músculos y dentro de los mismos. Estos gradientes de partida influyen en los gradientes finales de elementos añadidos durante el procesado, como la sal. La clasificación de filetes de salmón ahumado en función de su contenido graso y sal fue conseguida con un coeficiente de determinación de 0,97 y un error de predicción de 0,56, trabajando en transmitancia (Segtnan et al., 2009).

Trabajos similares se han realizado en reflectancia para carne de porcino durante su salado (Liu et al., 2013). De igual modo, es posible determinar el contenido de grasa intramuscular en cerdo en el rango espectral de 1193 a 1217 nm. Kamruzzaman et al.

(2012), obtuvieron modelos de regresión PLS para predecir la humedad, la grasa y la proteína en cordero fresco, con coeficientes de regresión que oscilaron entre 0,63 para la proteína y 0,88 para la humedad, y un error estándar de predicción SEP de 0,51 % para la humedad. ElMasry et al. (2013) también usó un modelo PLS para medir la humedad, la grasa y la proteína en ternera, obteniendo un coeficiente de determinación de 0,89 y un SEP de 0,46 % para humedad. Iqbal et al. (2013) predijeron la humedad, el color y el pH en rodajas de pavo cocido, obteniendo similares coeficientes de regresión, aunque con valores de SEP ligeramente superiores a los obtenidos por Elmasry et al. (2013). En cerdo, Barbin et al. (2013) y Talensa et al. (2013) predijeron la composición química (proteína, humedad y grasa) con éxito similar al de los anteriores. Liu et al (2013) prediciendo la humedad y contenido en sal en distintos cortes de cerdo fresco en diferente etapas de su proceso de salado, obtuvieron coeficientes de determinación de 0,93 y 0,91 y valores del SEP de 0,682 para la humedad y 0,007 para la sal, usando un instrumento que abarcaba el rango de 400-1000 nm. Gou et al. (2013), usando un rango más limitado predijo la humedad y la sal, obteniendo coeficientes de determinación de 0,96 y 0,91 y SEP de 1,34 % y 0,71, respectivamente.

El caso de la sal es especial, debido a sus implicaciones normativas y de salud. Los niveles finales de sal en jamón curado varían considerablemente desde un 4 % a un 10 % (Arnau et al., 1994) en el mismo músculo y entre músculos (dependiendo de la localización, el veteado, la permeabilidad a la humedad y a la sal de los tejidos que cubren el músculo, entre otros factores). Esta variabilidad debe tenerse en cuenta en el proceso de disminución de la sal de sus productos en el que se encuentran inmersos sus industriales, ya que la distribución irregular de sal o su modificación con respecto a las cantidades tradicionales afectan decisivamente a su calidad organoléptica (García-Gil et al., 2014). La evaluación del potencial de los sensores hiperespectrales en la industria del jamón ibérico para la categorización de sus productos por tipo de alimentación y contenido en sal, a la vez el estudio de la distribución de la misma en los distintos músculos del jamón tras su curado, fue el segundo objetivo de la presente tesis [*“Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”*, *Journal of Food Engineering* 153 (2015) 117-123]. Dentro del mismo objetivo, y como trabajo preliminar, se estudió la posibilidad de analizar el veteado de caña de lomo ibérica usando imagen hiperespectral [*“Methodology for marbling characterisation of pure Iberian cured loin*

base on hyperspectral imaging". NIR2013 Proceedings, 2-7 June, La Grande-Motte, France. Bellon-Maurel V., Williams P., Downey G., Eds]

Otros estudios han demostrado la capacidad de la tecnología hiperespectral para la identificación y autenticación de diferentes especies de carne roja. Por ejemplo, existe un trabajo que alcanzó unos porcentajes de acierto del 99,93 y 97 % en la clasificación de ternera, cerdo y cordero a partir del músculo *longissimus dorsi* (Kamruzzaman et al., 2012). Este tipo de trabajos también se han llevado a cabo con carne picada para la detección de adulteraciones (Kamruzzaman et al., 2013).

La detección de tumores en pollo mayores a 3 mm, de difícil detección a simple vista o con los equipos ópticos tradicionales, ha sido posible gracias a equipos hiperespectrales de fluorescencia que trabajan entre los 425 y los 711 nm (Kim et al., 2004; Kim et al., 2006; Kong et al., 2004).

La contaminación por de microbios puede ser determinada con altos niveles de precisión ($R^2 = 0,94$) mediante equipos de Imagen Hiperespectral en reflectancia (400-1000 nm), tanto en cerdo (Wang et al., 2010), como en ternera (Peng et al., 2011; Peng et al., 2009). Se han desarrollado métodos de detección de contaminaciones por *Escherichia coli* en cerdo en el mismo rango espectral (Tao et al., 2012).

Otro tipo de contaminaciones estudiadas son las fecales. Trabajando entre los 400 y los 1000 nm se han obtenido porcentajes de acierto del 90 % detectando contaminación fecal en pollo (Heitschmidt et al., 2007; Park et al., 2002; Park et al., 2007).

Otros productos de origen animal son las harinas o subproductos proteicos de origen animal. Se trata de un producto en el que la precisión en la determinación de la composición es de gran importancia por la implicación de la contaminación con harinas de vacuno en la difusión de la Encefalopatía Espongiforme Bovina (EEB), que derivó en la prohibición de su uso en piensos de animales de producción en la UE (Regulation (EC) No 1069/2009 of the European Parliament and of the Council of 21 October 2009), tal y como se recoge en los objetivos 1 y 2 de la presente tesis doctoral.

Riccioli et al. (2011) propusieron un método basado en imágenes hiperespectrales para distinguir entre harinas derivadas de pescado y animales terrestres usando píxeles representativos, alcanzando porcentajes de acierto del 99 %. Riccioli et al (2012) compararon el rendimiento de la Espectroscopía NIRS clásica con la Imagen

Hiperespectral en la detección y cuantificación de harina de ruminante en harinas de origen animal. Los resultados mostraron que si bien la Imagen Hiperespectral tiene un mayor potencial en términos de detección de contaminaciones debido a la información cuantitativa que facilita, la cuantificación con NIRS clásico arrojó mayores precisiones. El tercer y cuarto objetivos se centraron en esta problemática asociada a las harinas de origen animal. En el tercero se evaluó el potencial de los sensores hiperespectrales para la cuantificación y detección de contaminaciones con harinas de rumiantes en harinas de carne [*“Differentiation of beef and fish meals in animal feeds using chemometric analytic models”*, *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)], siendo el objeto del cuarto su evaluación para la realización de propuestas de optimización para identificación de la especie animal en harinas de carne [*“Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”*, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99].

En la Tabla 2, adaptada y completada a partir de Siche et al. (2015) se hace un resumen de dichas aplicaciones, indicando el producto, los parámetros medidos, el rango espectral, el coeficiente de determinación y el porcentaje de correctamente clasificados.

Tabla 2. Aplicaciones de equipos hiperespectrales en productos de origen animal. Se indica el producto, el objeto del análisis, el rango espectral, la precisión de la predicción o clasificación y la referencia.

<i>Producto</i>	<i>Parámetro</i>	<i>λ (nm)</i>	<i>Precisión (R^2 o %)</i>	<i>Referencia</i>
Ternera	Terneza	496-1036	0,67	Cluff et al., 2008.
	Terneza	400-1000	96,4 %	Naganathan et al., 2008a.
	Terneza	900-1700	77 %	Naganathan et al., 2008b
	Terneza	450-900	92,9 %	Naganathan et al., 2015
	Terneza	400-1100	0,94	Peng y Wu 2008
	Autenticación	900-1700	99 %	Kamruzzaman et al., 2012
	Pigmentos	328-1115	0,95	Xiong et al., 2015a
	Conteo de bacterias viables	400-1100	0,96	Peng et al., 2009

	Conteo de bacterias viables	400-1000	0,95	Peng et al., 2011
Cerdo	Calidad	400-1000	85 %	Qiao et al., 2007a
	Calidad	430-980	87,5 %	Qiao et al., 2007b
	Autenticación	900-1700	93 %	Kamruzzaman et al., 2012
	Frescura	900-1700	98 %	Barbin et al., 2013
	Pérdida de humedad	430-980	0,77	Qiao et al., 2007
	pH	430-980	0,55	Qiao et al., 2007
	Color	430-980	0,86	Qiao et al., 2007
	pH en carne salada	400-1000	0,79	Liu et al., 2014a
	Humedad en carne salada	400-1000	0,92%	Liu et al., 2014b
	Conteo de microbios viables	900-1700	0,82	Barbin et al., 2012
	Terneza	400-1100	0,93	Tao et al., 2012
	E. coli	400-1100	0,88	Tao et al., 2012
Pollo	Presencia de huesos	400-1000	100 %	Yoon et al., 2008
	Fragmentos de hueso	400-1000	100 %, falsos positivos 10 %	Yoon et al., 2008
	Salubridad	400-1000	99 % salubres y 96 % insalubres	Chao et al., 2008
	Tumores en piel	447-951	80 %	Nakariyakul et al., 2009
	Tumores	425-711	76 %	Kim et al., 2004
	Tumores	425-711	82 %	Kong et al., 2004
	Tumores	425-711	98 %	Kim et al., 2006
	Contaminación fecal	400-1000	99 %	Heitschmidt et al., 2007
	Contaminación fecal	430-900	96 %	Park et al., 2006
	Contaminación fecal	400-900	97 %	Park et al., 2002
	Contaminación fecal	400-900	90 %	Park et al., 2007

	Enfermedades	400-900	93.5 %	Yang et al., 2005
	Discriminación entre carne de pollo criado en libertad y en jaula	328-1115	93 %	Xiong et al., 2015
Cordero	Clasificación de músculos	900-1700	100 %	Kamruzzaman et al., 2011
	Autenticación	900-1700	97 %	Kamruzzaman et al., 2012
	Adulteración	900-1700	0,98	Kamruzzaman et al., 2013
Pescados	Frescura	892-2495	0,59	Siripatrawan et al., 2011
	Humedad	760-1040	0,92 (RMSECV = 0,7 %)	Wold et al., 2006
	Humedad	760-1040	0,94	El Masry et al., 2008
	Grasa	760-1040	0,91	El Masry et al., 2008
	Oxidación lipídica	400-1100	0,83	Cheng et al., 2015a
	Microorganismos viables	400-1000	0,90	Cheng et al., 2015b
	Microorganismos viables	400-1700	0,99	Wu et al., 2013
	Parásitos	350-950	100 %	Heia et al. 2007.
	Frescura	400-1000	95 – 98 %	Dai et al., 2015
Harinas de origen animal	Composición	1100-1600	83 -99 según especie (riesgo de falsos positivos)	Riccioli et al., 2012
	Composición	1100-1600	99 %	Riccioli et al., 2011

4.4. Bibliografia.

- Amigo, J. M., Cruz, J., Bautista, M. , MasPOCH, S., Coello, J., Blanco, M. 2008. Study of pharmaceutical samples by NIR chemical-image and multivariate analysis. Trends in Analytical Chemistry, Vol. 27, N° 8.
- Amigo, J. M., Ravn, C. 2009. Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging. European Journal of Pharmaceutical Sciences. 37 (2009) 76-82
- Ariana D. P., Lu R., Guyer D. E. 2006. Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. Comput Electron Agric 53:60–70
- Ariana DP, Lu R. 2008. Detection of internal defect in pickling cucumbers using hyperspectral transmittance imaging. Trans ASABE 51:705–713
- Arnau, J., Guerrero, L., Casademont, G., Gou, P. 1994. Physical and chemical changes in different zones of normal and PSE dry cured ham during processing. Food Chem. 52, 62-69
- Barbin D., ElMasry G., Sun D-W., Allen P., Noha M. 2012 Nondestructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging. Innov. Food. Sci. Emerg. Technol. 17:180–191
- Barbin D., Sun D.-W., Su C. 2013. NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen–thawed porcine *longissimus dorsi* muscles. Innov. Food Sci. Emerg. Techol. 18:226–236, 92, 60
- Benson, I.B. 2003. Near infrared absorption technology for analysing food composition. In Food Authenticity and Traceability. Lees, M. (Ed.). Woodhead Publishing Ltd, Cambridge, England. pp. 101-130.
- Bock J. E., Connelly R. K. 2008. Innovative Uses of Near-Infrared Spectroscopy in Food Processing. Journal of Food Science, Vol. 73, Nr. 7, 200

- Burger, J. y Geladi, P. 2006. Hyperspectral NIR image regression part II: dataset preprocessing diagnostics. 2006. J. Chemom. 20, 106-119.
- Buttingsrud, B. and B.K. Alsberg, 2006. Super-resolution of hyperspectral images. Chemometrics Intell. Lab. Syst., 84: 62-68.
- Ciurezak, E.W. 2001. Principles of Near-Infrared Spectroscopy. In Handbook of Near Infrared Analysis. 2nd edition. Burns, D.A., Ciurezak E.W. (Eds.). Practical Spectroscopy Series, Vol. 27. Marcel Dekker, New York. pp. 7-18
- Cluff K., Naganathan G. K., Subbiah J., Lu R., Calkins C. R., Samal A. 2008. Optical scattering in beef steak to predict tenderness using hyperspectral imaging in the VIS-NIR region. Sens Instrum. Food Qual Saf 2:189–196
- Chao, K., Yang, C.-C., Kim, M.S., Chan, D.E. .2008. High throughput spectral imaging system for wholesomeness inspection of chicken. Applied Engineering in Agriculture 24 (4), 475-485.
- Chen S, Zhang F, Ning J, Liu X, Zhang Z, Yang S. 2015. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. Food Chem 172:788–793
- Cheng J-H., Sun D-W. 2015b. Rapid and non-invasive detection of fish microbial spoilage by visible and near infrared hyperspectral imaging and multivariate analysis. LWT Food. Sci. Technol. 62(2):1060–1068
- Cheng J-H., Sun D-W., Pu H-B., Wang Q-J., Chen Y-N. 2015a. Suitability of hyperspectral imaging for rapid evaluation of thiobarbituric acid (TBA) value in grass carp (*Ctenopharyngodon idella*) fillet. Food. Chem. 171:258–265
- Cheng X, Chen YR, Tao Y, Wang CY, Kim MS, Lefcourt AM .2004. A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. Trans ASAE 47(4):1313–1320
- Dahm, D.J., y Dahm, K.D. 2001. The physics of near infrared scattering. En P. Williams y K. Norris (Eds.) Near Infrared Technology in the Agricultural and Food Industries (2º Edición), (p. 1-18). America Association of Cereal Chemists, Inc. St Paul, Minnesota, USA.

- Dai Q., Cheng J-H., Sun D-W. 2015. Potential of visible/nearinfrared hyperspectral imaging for rapid detection of freshness in unfrozen and frozen prawns. *J. Food. Eng.* 149:97–104
- de Juan, A., Tauler, R., Dyson, R., Marcolli, C., Rault, M., Maeder, M. 2004. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *Trac-Trends in Analytical Chemistry* 23 (1), 70–79.
- Ding, Y., Li, M. , An, D. , Li, S. 2011. Prediction of chlorophyll content using spectral response characteristics of greenhouse tomato. *Transactions of the Chinese Society of Agricultural Engineering*. Volume 27, 5, 244-247.
- El Masry G., Wold J. P. 2008. High-speed assessment of fat and water content distribution in fish fillets using online imaging spectroscopy. *J. Agric. Food. Chem* 56(17):7672–7677
- ElMasry G, Kamruzzaman M, Sun D-W, Allen P. 2012. Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Crit Rev Food Sci* 52(11):999–1023
- ElMasry G, Wang N, ElSayed A, Ngadi M. 2007. Hyperspectral imaging for non-destructive determination of some quality attributes for strawberry. *J Food Eng* 81:98–107
- ElMasry G, Wang N, Vigneault C. 2009. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biol Technol* 52(1):1–8
- ElMasry G, Wang N, Vigneault C, Qiao J, ElSayed A. 2008. Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT Food Sci Technol* 41:337–345.
- ElMasry G., Wang N., Vigneault C. 2009. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biol. Technol.* 52(1):1–8

- Elmasry, G., Sun, D.W., Alen, P., 2013. Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. *J. Food Eng.* 117 (2), 235-246
- Fearn, T., Riccioli, C., Garrido-Varo, A. and Guerrero-Ginel, E. 2009. On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems* 96, pp. 22-26.
- Fernández Cabanas, V. 2003 Métodos de procesamiento de la señal espectroscópica NIRS: Aplicación al análisis cuantitativo y cualitativo de productos agroalimentarios, Tesis Doctoral, Universidad de Córdoba
- García-Gil, N., Muñoz, I., Santos-Garcés, E., Arnau, J., Gou, P. 2014, Salta uptake and moisture loss in hams with different moisture contents at the lean surface and at different salting temperatures. *Meat. Sci.* 96 (1), 65-72
- Garrido, A & De-Pedro Sanz, Emiliano. (2007). The role of near infrared spectroscopy in verifying label information in agro-forestry. *Handbook of near-infrared spectroscopy analysis. Practical spectroscopy series.* 387-398.
- Geladi P, Burger J, Lestander T (2004) Hyperspectral imaging: calibration problems and solutions. *Chemometrics Intellig Lab Syst* 72(2):209–217
- Geladi P., McDougall, D., Martens, H. 1985 Linearization and Scatter Correction for Near-Infrared Reflectance Spectra of Meat. *Appl. Spectrosc.* 39, 491-500.
- Goel, P.K.; Prasher, S.O.; Patel, R.M.; Landry, J.A.; Bonnell, R.B.; Viau, A.A. 2003. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. *Comput. Electron. Agric.*, 39, 67-93.
- Gómez-Sanchis, J., Gómez-Chova, L., Aleixos, N., Camps-Valls, G., Montesinos Herrero, C., Moltó, E., Blasco, J. 2008. Hyperspectral system for early detection of rottenness caused by *Penicillium digitatum* in mandarins. *Journal of Food Engineering* 89 (1), 80-86.
- Gou, P., Santos-Garcés, E., Hoy, M., Wold, J. P., Liland, K. H., Fulladosa, E., 2013. Feasibility of NIR interactance hyperspectral imaging for on-line measurement of

crude composition in vacuum packed dry-cured ham slices. *Meat Sci.* 95 (2), 250-244.]

Govender, M., Chetty, K., Bulcock, H. 2007. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA*. Vol. 33. Nº. 2 April 2007.

Gowen A. A., Taghizadeh M., O'Donnell C. P. 2009. Identification of mushrooms subjected to freeze damage using hyperspectral imaging. *J. Food. Eng.* 93(1):7–12

Gowen, A.A., O'Donnell, C. P., Taghizadeh M., Gaston E., O'Gorman A., Cullen P. J., Frias J. M., Esquerre C., Downey G. 2008. Hyperspectral imaging for the investigation of quality deterioration in sliced mushrooms (*Agaricus bisporus*) during storage. *Sensing and Instrumentation for Food Quality and Safety*, 2:133–143.

Haralick, R. M., Shanmugam, K., Dinstein, I. H. 1973. Textural features for image classification, *IEEE Trans. Syst. Man. Cybern.* , 6 (1973), 610–621

Hart, J.R., y Norris, K.H. 1962. Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. *Cereal Chemistry*, 39, 94-99.

Heia, K., Sivertsen, A.H., Stormo, S.K., Elvevoll, E., Wold, J.P., Nilsen, H. 2007. Detection of nematodes in cod (*Gadus morhua*) fillets by imaging spectroscopy. *Journal of Food Science* 72 (1), E11-E15.

Heitschmidt G. W., Park B., Lawrence K. C., Windham W. R., Smith D. P. 2007 Improved hyperspectral imaging system for fecal detection on poultry carcasses. *Trans ASABE* 50(4):1427–1432

Herschel, F.W. 1800. Investigation of the powers of the prismatic colours to heat and illuminate objects. *Philosophical Transactions of the Royal Society of London*, 90, 255-329.

Hong, T., Qiao, J., Ning, W., Ngadi, M.O., Zhao, Z., Li, Z. 2007. Non destructive inspection of Chinese pear quality based on hyperspectral imaging technique

- Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering, 23 (2), 151-155.
- Hotelling, H. 1947. Multivariate Quality Control, in C. Eisenhart, M. W. Hastay and W. A. Wallis (Eds.); Techniques of Statistical Analysis, McGrawHill, New York
- Huang, M., Lu, R. 2010. Apple mealiness detection using hyperspectral scattering technique. *Postharvest Biology and Technology* 58 (3), 168-175.
- Huang, X., Xiaobo, L., Zhang, L. 2014. A multichannel Gray Level Co-occurrence Matrix for Multi/Hyperspectral Image texture representation, *Remote Sens.*, 6, 8424-8445.
- Iqbal, A., Sun, D. W., Allen, P. 2013. Prediction of moisture, color and pH in cooked, pre-sliced turkey hams by NIR hyperspectral imaging system. *J. Food Eng.* 117 (1), 42-51
- Janesick J. R. 2001. Scientific Charge-Coupled Devices. SPIE Press Monograph Vol. PM83, 1st Edition
- Jiang, L.; Zhu, B.; Rao, X.; Berney, G.; Tao, Y. 2007. Discrimination of black walnut shell and pulp in hyperspectral fluorescence imagery using Gaussian kernel function approach. *J. Food Eng.*, 81, 108-117.
- Jianwei Qin, Moon S. Kim, Duanglin Chao, Diane E., Chan, Stephen R. Delwiche y Byoung-Kwan Cho. 2017. Line-Scan Hyperspectral Imaging Techniques for Food Safety and quality Applications. *Applied Sciences*, 7, 125
- Kamruzzaman M., Barbin D., ElMasry G., Sun D-W., Allen .P. 2012. Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innov Food. Sci. Emerg. Techol* .16:316–325
- Kamruzzaman M., El Masry G., Sun D-W., Allen P. 2011. Application of NIR hyperspectral imaging for discrimination of lamb muscles. *J. Food. Eng.* 104:332–340
- Kamruzzaman M., Sun D-W, ElMasry G., Allen P. 2013 Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta* 103:130–136.

- Kamruzzaman, M., ElMasry, G., Sun, D. W., Allen, P. 2012. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innovative Food Sci. Emerging Technol.* 16, 218-226
- Kennard, R.W. and Stone, L.A. 1969. Computer aided design of experiments. *Technometrics* 11, 137-148.
- Kim I. , Xu C., Kim M. S. 2006. Poultry skin tumor detection in hyperspectral images using radial basis probabilistic neural network. *Adv. Neural. Netw.* 3973:770–776
- Kim I., Kim M. S., Chen Y. R., Kong S. G. 2004. Detection of skin tumors on chicken carcasses using hyperspectral fluorescence imaging. *Trans ASAE* 47(5):1785–1792
- Kim I., Kim M. S., Chen Y. R., Kong S. G. 2004. Detection of skin tumors on chicken carcasses using hyperspectral fluorescence imaging. *Trans ASAE* 47(5):1785–1792
- Kim M. S., Chen Y.R., Cho B. K., Chao K., Yang C. C., Lefcourt A. M., Chan D. 2007. Hyperspectral reflectance and fluorescence linescan imaging for online defects and fecal contamination inspection of apples. *Sens. Instrum. Food. Qual. Saf.* 1:151–159
- Kim, M.S.; Chen, Y.; Mehl, P.M. 2001. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE* 2001, 44, 721–729.
- Kong S. G., Chen Y. R., Kim I, Kim MS. 2004. Analysis of hyperspectral fluorescence images for poultry skin tumor inspection. *Appl. Opt.* 43(4):824–83
- Lachenal, G. 2000. Introduction à la spectroscopie infrarouge. In *La Spectroscopie Infrarouge et Ses Applications Analytiques*. Bertand, D., Dufuor, E. (Eds.). TEC & DOC Editions, Paris, Francia. pp. 31-75.
- Landgrebe D. 1999. On information extraction principles for hyperspectral data. *Cybernetics* 28 part c, 1, 1-7

- Lee H, Kim MS, Jeong D, Delwiche SR, Chao K, Cho B-K. 2014. Detection of cracks on tomatoes using a hyperspectral near-infrared reflectance imaging system. *Sensors* 14:18837–18850
- Lee, D. J., Schoenberger, R., Archibald, J., McCollom, S. 2008. Development of a machine vision system for automatic date grading using digital reflectance near-infrared imaging. *Journal of Food Engineering* Volume 86, Issue 3, 388398
- Liu D, Sun D-W, Qu J, Zeng X-A, Pu H, Ma J. 2014b. Feasibility of using hyperspectral imaging to predict moisture content of porcine meat during salting process. *Food Chem* 152:197–204
- Liu D., Pu H., Sun D-W., Wang L., Zeng X-A. 2014a. Combination of spectra and texture data of hyperspectral imaging for prediction of pH in salted meat. *Food Chem* 160:330–337
- Liu M., Zhang L., Guo E. 2008. Hyperspectral laser-induced fluorescence imaging for non-destructive assessing soluble solids content of orange. *Comput. Comput. Technol. Agric.* 1:51–59
- Liu Y, Chen YR, Wang CY, Chan DE, Kim MS. 2005. Development of a simple algorithm for the detection of chilling injury in cucumbers from visible/near-infrared hyperspectral imaging. *Appl Spectrosc* 59(1):78–85
- Liu, D., Qu, J., Sun, D.W., Pu, H., Zeng, X. A., 2013. Non-destructive prediction of salt contents and moisture activity of porcine meat slices by hyperspectral imaging in a salting process. *Innovative Food Sci. Emerging Technol.* 20, 316-323
- Liu, Y., Chen, R., Wang, C. Y., Chan, D. E., Kim, M. S. 2006. Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers; spectral and data analysis. *Applied Engineering in Agriculture*. Vol. 22(1): 101-111
- Liu, Y., Chen, Y.-R., Kim, M.S., Chan, D.E., Lefcourt, A.M. 2007. Development of simple algorithms for the detection of fecal contaminants on apples from visible/near infrared hyperspectral reflectance imaging. *Journal of Food Engineering* 81 (2), 412-418.

- Lu R 2003. Detection of bruises on apple using near-infrared hyperspectral imaging. Trans ASAE 46(2):523–530
- Lu R, Peng Y. 2006. Hyperspectral scattering for assessing peach fruit firmness. Biosyst Eng 93(2):161–171
- Lu, R. 2007. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images. Sens. & Instrumen. Food Qual. (2007) 1:19–27
- Lu, R.; Chen, Y. 1999. Hyperspectral imaging for safety inspection of food and agricultural products. Proc. SPIE 1999, 3544
- Mandelbrot, B. 1980. Fractal aspects of the iteration of $z \rightarrow \lambda z(1-z)$ for complex λ , z . Annals of the New York Academy of Sciences, 357, 249-259.
- Manne, R. 1987. Analysis of two partial-least-squares algorithms for multivariate calibration. Chemometrics and Intelligent Laboratory Systems, 2, 187-197.
- Martens, H., y Næs, T. 1985. Multivariate calibration by data compression. En H.A. Martens (Ed.), Multivariate Calibration. Quantitative Interpretation of Non-selective Chemical Data, (p. 167-286. Tesis doctoral, Technical University of Norway, Trondheim.
- Massart, D.L Vandeginste, B.G.M Buydens, L.M.C. De Jong, S.P. Lewi J. and Smeyers-Verbeke, S. 1997 In: B.G.M Vandeginste and S.C Rutan, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier pp. 1-20.
- Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., De Jong, S., Lwei, P.J., y Smeyers-Verbeke, J. 1997. Principal Component. En D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lwei, y J. SmeyersVerbeke (Eds.), Handbook of Chemometrics and Qualimetrics. Part B. (p. 519-556). Elsevier Science, Amsterdam, The Netherlands.
- McCann, M.C.; Defernez, M.; Urbanowicz, B.R.; Tewari, J.C.; Langewisch, T.; Olek, A.; Wells, B.; Wilson, R.H.; Carpita, N.C. 2007. Neural network analyses of infrared spectra for classifying cell wall architectures. Plant Physiol., 143, 1314-1326.

- McClure, W. F. 2003. 204 Years of near Infrared Technology: 1800–2003. *J. Near Infrared Spectrosc.* 11, 487-518
- Menesatti P., Zanella A. S., D’Andrea S., Costa C., Paglia G., Pallottino F. 2009. Supervised multivariate analysis of hyperspectral NIR images to evaluate the starch index of apples. *Food Bioprocess Technol* 2:308–314
- Miller, C.E. 2001. Chemical principles of near infrared spectroscopy. In *Near Infrared Technology in the Agricultural and Food Industries*. 2nd edition. Williams, P.K., Norris, K. (Eds.). American Association of Cereal Chemists, Inc. St Paul, Minnesota, USA. pp. 19-37.
- Morata A., Loira I., Vejarano R., Bañuelos M. A., Sanz P., Otero L., Suárez-Lepe J. A. 2015. Grape processing by high hydrostatic pressure: effect on microbial populations, phenol extraction and wine quality. *Food Bioprocess. Technol.* 8(2):277–286
- Naganathan G. K., Cluff K., Samal A., Calkins C. R., Jones D. D., Lorenzen C. L., Subbiah J. 2015. A prototype on-line AOTF hyperspectral image acquisition system for tenderness assessment of beef carcasses. *J Food Eng* 154:1–9
- Naganathan G. K., Grimes L. M., Subbiah J., Calkins C. R., Samal A., Meyer G. E. 2008. Visible/near-infraredhyperspectralimagingfor beef tenderness prediction. *Comput. Electron. Agric.* 64:225–233
- Naganathan G. K., Grimes L. M., Subbiah J., Calkins C. R., Samal A, Meyer G. E. 2008a.Visible/near-infraredhyperspectralimagingfor beef tenderness prediction. *Comput. Electron. Agric.* 64:225–233
- Naganathan G. K., Grimes L. M., Subbiah J., Calkins C. R., Samal A., Meyer G. E. 2008b. Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction. *Sens. Instrum. Food Qua.l Saf* 2:178–188
- Nagata M, Tallada JG, Kobayashi T. 2006. Bruise detection using NIR hyperspectral imaging for strawberry (*Fragaria x ananassa* Duch.). *Environ Control Biol* 44(2):133–142

- Nagata M., Tallada J. G., Kobayashi T. 2006. Bruise detection using NIR hyperspectral imaging for strawberry (*Fragaria x ananassa* Duch.). *Environ Control Biol* 44(2):133–142
- Nakariyakul, S., and Casasent, D.P. 2009. Fast feature selection algorithm for poultry skin tumor detection in hyperspectral data. *Journal of Food Engineering* 94, 358–365
- Nguyen Do Trong NN, Tsuta M, Nicolai B.M., Baerdemaeker JD, Saeys W. 2011. Prediction of optimal cooking time for boiled potatoes by hyperspectral imaging. *J Food Eng* 105(4):617–624
- Nicolai, B.M., Lötze, E., Peirs, A., Scheerlinck, N., Theron, K.I. 2006. Nondestructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. *Postharvest Biology and Technology* 40 (1), 1-6.
- Nogales-Bueno J, Hernández-Hierro J. M., Rodríguez-Pulido F. J., Heredia F. J. 2014. Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: a preliminary approach. *Food Chem* 152:586–591
- Nogales-Bueno J, Rodríguez-Pulido FJ, Heredia FJ, Hernández-Hierro JM. 2015. Comparative study on the use of anthocyanin profile, color image analysis and near-infrared hyperspectral imaging as tools to discriminate between four autochthonous red grape cultivars from La Rioja (Spain). *Talanta* 131:412–416
- Nogales-Bueno J., Baca-Bocanegra B., Rodríguez-Pulido F. J., Heredia F. J., Hernández-Hierro J. M.. 2015. Use of near infrared hyperspectral tools for the screening of extractable polyphenols in red grape skins. *Food Chem* 172:559–564.
- Noh H. K., Lu R. 2007. Hyperspectral laser-induced fluorescence imaging for assessing apple fruit quality. *Postharvest Biol. Technol.* 43:193–201
- Okamoto, H., Lee, W.S. 2009. Green citrus detection using hyperspectral imaging. *Computers and Electronics in Agriculture* 66 (2), 201-208.

- Park B. , Lawrence K. C., Windham W. R., Buhr R. J. 2002 Hyperspectral Imaging for detecting fecal and ingesta contaminants on poultry carcasses. *Trans ASAE* 45(6):2017–2026
- Park B., Lawrence K. C., Windham W. R., Smith DP. 2006. Performance of hyperspectral imaging system for poultry surface fecal contaminant detection. *J. Food. Eng.* 75:340–348
- Park B., Windham W. R., Lawrence K. C., Smith DP. 2007. Contaminant classification of poultry hyperspectral imagery using spectral angle mapper algorithm. *Biosyst Eng* 96(3):323–333
- Peng Y, Lu R. 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biol Technol* 48:52–62
- Peng Y. , Zhang J. , Wang W. , Li Y, Wu J, Huang H, Gao X., Jiang W. 2011. Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *J Food. Eng* 102:163–169
- Peng Y., Lu R. 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biol. Technol.* 48:52–62
- Peng Y., Wu J. 2008. Hyperspectral scattering profiles for prediction of beef tenderness. American Society of Agricultural and Biological Engineers, St. Joseph, Michigan
- Peng Y., Zhang J, Wu J., Hang H. 2009. Hyperspectral scattering profiles for prediction of the microbial spoilage of beef. *Proc SPIE* 7315:73150Q1–73150Q12
- Peng Y., Zhang J., Wang W., Li Y., Wu J., Huang H., Gao X., Jiang W. 2011 Potential prediction of the microbial spoilage of beef using spatially resolved hyperspectral scattering profiles. *J. Food. Eng.* 102:163–169
- Peng, Y., Lu, R. 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biology and Technology* 48 (1), 52-62

- Pierna, J.A.F.; Baeten, V., Renier, A.M., Cogdill, R.A, Dardenne, P. 2004. Combination of support vector machines (SVM) and near infrared (NIR) imaging spectroscopy for the detection of meat and bonemeal (MBM) in compound feeds. *Chemometrics.*, 18, 341-349.
- Qiao J., Ngadi M. O., Wang N, Gunenc A, Monroy M, Gariepy C, Prasher S. 2007b. Pork quality classification using a hyperspectral imaging system and neural network. *Int J Food Eng* 3(1):1–12
- Qiao J., Ngadi M. O., Wang N, Gariepy C, Prasher SO. 2007a. Pork quality and marbling level assessment using a hyperspectral imaging system. *J Food. Eng.* 83(1):10–16
- Qiao J., Wang N., Ngadi M., Gunenc A., Monroy M, Gariepy C, Prasher S. 2007. Prediction of drip-loss, pH and color for pork using a hyperspectral imaging technique. *Meat Sci.* 76(1):1–8
- Qin J, Lu R. 2008. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biol Technol* 49:355–365
- Qin J, Lu R, Peng Y. 2009. Prediction of apple internal quality using spectral absorption and scattering properties. *Trans ASABE* 52(2):499–507
- Qin J., Lu R. 2008. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest. Biol. Technol.* 49:355–365
- Qing, Z., Zude, M. 2007. Wavelength selection for predicting physicochemical properties of apple fruit based on near-infrared spectroscopy. *J. Food Qual.* 30, 511-526.
- Quinlan, J. R. 1986. Induction of decision trees, *Machine Learning*, 1, 1 (1986), 81-106
- Rajkumar P, Wang N, Eimasry G, Raghavan GSV, Gariepy Y 2012. Studies on banana fruit quality and maturity stages using hyperspectral imaging. *J Food Eng* 108:194–200

- Riccioli, C. 2011. Detección y cuantificación de la especie en harinas proteicas de origen animal mediante el uso de sensores hiperespectrales. Tesis Doctoral. Universidad de Córdoba.
- Riccioli, C., Pérez-Marín, D. C., Guerrero-Ginel, J. E., Wouter, S., Garrido-Varo, A.. 2011. Pixel selection for near-infrared chemical imaging (NIR-CI) discrimination between fish and terrestrial animal species in animal protein by-products meals. *Appl Spectrosc.*, 65, 7 (2011), 771-81.
- Riccioli, D. C. Pérez-Marín, J. E. Guerrero-Ginel, T. Fearn, A. Garrido-Varo. 2012. Detection and quantification of ruminant meal in processed animal proteins: a comparative study of near infrared spectroscopy and near infrared chemical imaging. *J. Near Infrared Spectrosc.*, 20, 6 (2012), 623-633.
- Segtnan V. H. , Høy M., Sørheim O., Kohler A., Lundby F., Wold J. P., Ofstad R. 2009. Non contact salt and fat distributional analysis in salted and smoked salmon fillets using X-ray computed tomography and NIR interactance imaging. *J. Agric. Food. Chem.* 57:1705–1710
- Shenk, J.S., & Westerhaus, M.O. 1995. Analysis of agriculture and food products by near infrared reflectance spectroscopy. Monograph NIRSystems.
- Shenk, J.S., Westerhaus, M.O. 1995. Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy. Monograph, NIRSystems, Inc., 12101 Tech Road, Silver Spring, MD 20904, USA.
- Shenk, J.S., y Westerhaus, M.O. 1991. Population definition, sample selection and calibration procedures for near infrared spectra and modified partial least squares regression. *Crop Science*, 31, 469-474.
- Shenk, J.S.; Workman, J.J., Westerhaus, M.O. 1992. Application of NIR Spectroscopy to agricultural products. In *Handbook of Near-Infrared Analysis*. Burns, D.A., Ciurczak, E.W. (Eds.). Vol. 13. Practical Spectroscopy Series, Marcel Dekker, New York. pp. 383-431.
- Siche. R., Vejarano, R., Aredo, V., Velasquez, L., Saldaña, E., Quevedo, R. 2015. Evaluation of Food Quality and Safety with Hyperspectral Imaging (HSI). *Food Eng. Rev.* DOI 10.1007/s12333-015-9137-8

- Siesler, H.W., Ozaki, Y., Kawata, S., y Heise, H.M. 2002. Near Infrared Spectroscopy: principles, instruments, applications. Wiley-Vch, Weinhen, Alemania.
- Siripatrawan U, Makino Y, Kawagoe Y, Oshita S. 2011. Rapid detection of Escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta* 85:276–281
- Siripatrawan U, Makino Y, Kawagoe Y, Oshita S. 2011. Rapid detection of Escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta* 85:276–281
- Staab, E.S., Slaughter, D.C., Zhang, Y., Giles, D.K. 2009. Hyperspectral imaging system for precision weed control in processing tomato. American Society of Agricultural and Biological Engineers Annual International Meeting 2009 Volume 7, 4644-4656.
- Sugiyama, T., Sugiyama, J., Tsuta, M., Fujita, K., Shibata, M., Kokawa, M., Araki, T., Nabetani, H., and Sagara, Y. 2010. NIR spectral imaging with discriminant analysis for detecting foreign materials among blueberries. *Journal of Food Engineering* 101, 3, 244-252.
- Suthiluk P, Saranwong S, Kawano S, Numthuam S, Satake T. 2008. Possibility of using near infrared spectroscopy for evaluation of bacterial contamination in shredded cabbage. *Int J Food Sci Technol* 43:160–165
- Talensa, P., Mora, L., Morsy, N., Douglas, Barbin, F., ElMasry, G., Sun, D. W. 2013. Prediction of moisture and protein contents and quality classification of Sarnish cooked ham using NIR hyperspectral imagin. *J. Food Eng.* (3), 272-280
- Tao F., Peng Y., Li Y., Chao K., Dhakal S. 2012. Simultaneous determination of tenderness and Escherichia coli contamination of pork using hyperspectral scattering technique. *Meat Sci.* 90:851–857
- Tao F., Peng Y., Li Y., Chao K., Dhakal S. 2012. Simultaneous determination of tenderness and Escherichia coli contamination of pork using hyperspectral scattering technique. *Meat Sci.* 90:851–857].

- Tsuta, M., Takao, T., Sugiyama, J., Wada, Y. and Sagara, Y. 2006. Foreign Substance Detection in Blueberry Fruits by Spectral Imaging. *Food Science and Technology Research* Vol. 12, 2 96-100.
- Vélez N, Gómez-Sanchis J, Chanona-Pérez J, Carrasco J. J., Millán-Giraldo M, Lorente D, Cubero S, Blasco J. 2014 Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning. *Biosyst Eng* 122:91–98
- Wang W., Li C., Tollner E. W., Gitaitis RD, Rains G. C. 2012. Shortwave infrared hyperspectral imaging for detecting sour skin (*Burkholderia cepacia*)-infected onions. *J Food Eng* 109(1):38–48
- Wang W., Peng Y. K., Zhang X. L. 2010. Study on modeling method of total viable count of fresh pork meat based on hyperspectral imaging system. *Spectrosc. Spectr. Anal.* 30(2):411–415, 7
- Wold, J.F., Johansen, I.-R., Haugholt, K.H., Tschudi, J., Thielemann, J., Segtnan, V.H., Narum, B., Wold, E. 2006. Non-contact transreflectance near infrared imaging for representative on-line sampling of dried salted codfish (bacalao). *Journal of Near Infrared Spectroscopy* 14 (1), 59-66
- Workman, J.J., Shenk, J. 2004. Understanding and using the near-infrared spectrum as an analytical method. In *Near-Infrared Spectroscopy in Agriculture*. Roberts C.A., Workman, J., Reeves III, J.B. (Eds.). ASA, CSSA y SSSA, Inc., Madison, Wisconsin, USA. pp. 3-10
- Wu D., Sun D-W. 2013. Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh. *Talanta* 111:39–46
- Xing J. ,Bravo C. ,Jancsok P. T., Ramon H., De Baerdemaeker J. 2005. Detecting bruises on “golden delicious” apples using hyperspectral imaging with multiple wavebands. *Biosyst Eng* 90(1):27–36
- Xing, J., Guyer, D., Ariana, D., Lu, R. 2008. Determining optimal wavebands using genetic algorithm for detection of internal insect infestation in tart cherry. *Sensing and Instrumentation for Food Quality and Safety* 2 (3), 161-167.

- Xiong Z., Sun D-W., Pu H., Zhu Z., Luo M. 2015. Combination of spectra and texture data of hyperspectral imaging for differentiating between free-range and broiler chicken meats. *LWT Food Sci. Technol.* 60(2):649–655
- Xiong Z., Sun D-W., Xie A., Pu H., Han Z., Luo M. 2015a. Quantitative determination of total pigments in red meats using hyperspectral imaging and multivariate analysis. *Food Chem.* 178:339–34
- Yang C.C., Chao K., Chen Y-R., Early H. L. 2005a. Systemically diseased chicken identification using multispectral images and region of interest analysis. *Comput Electron Agric* 49:255–271
- Yoon S. C., Lawrence K. C., Smith D. P., Park B., Windham W. R. 2008. Embedded bone fragment detection in chicken fillets using transmittance image enhancement and hyperspectral reflectance imaging. *Sens Instrum Food Qual. Saf.* 2:197–207
- Zamora Rojas, E. 2013. Control no destructivo e in situ de productos y procesos en la industria del cerdo ibérico usando sensores espectrales de infrarrojo cercano. Tesis doctoral. Universidad de Córdoba.
- Zhao J, Vittayapadung S, Chen Q, Chaitep S, Chuaviroj R. 2009. Non-destructive measurement of sugar content of apple using hyperspectral imaging technique. *Maejo Int J Sci Technol* 3(1):130–142
- Zhao J., Vittayapadung S., Chen Q., Chaitep S., Chuaviroj R. 2009. Non-destructive measurement of sugar content of apple using hyperspectral imaging technique. *Maejo Int. J. Sci. Technol.* 3(1):130–142
- Zhao, X., Burks, T.F., Qin, J., Ritenour, M.A. 2010. Effect of fruit harvest time on citrus canker detection using hyperspectral reflectance imaging. *Sensing and Instrumentation for Food Quality and Safety* 4 (3), 126-135.

5. RESULTADOS Y DISCUSIÓN

5.1.-Puesta a punto del equipo hiperespectral.

Los equipos comerciales, cada vez más usados en evaluaciones de rutina en la industria agrolimentaria, ofrecen soluciones completas para aplicaciones definidas. Los equipos que usan soluciones modulares, por el contrario, ofrecen una mayor flexibilidad y versatilidad ya que cada uno de sus componentes y sus configuraciones puede ser modificado y optimizado para distintas aplicaciones. A continuación se expone la comparación del equipo de barrido lineal modular o no comercial con un equipo comercial de laboratorio.

El equipo de barrido lineal, usado para la consecución de los cuatro objetivos de la presente tesis doctoral (Fig. 7), consistió en un espectrómetro (ImSpector V10E) que cubre el rango NIR (900-1700 nm) con una resolución espectral de 5nm, una cámara (320 x 256) FPA-1.7 XEVA- Xenics-320-CCD LS y 1 μ min-5 min de tiempo de exposición, relación focal de F/1.4 y conexión USB. Se usaron dos lámparas halógenas de 250 w. El barrido se realizó gracias a una plataforma controlada por un motor de pasos coordinada con la cámara (VXM, Velmex Inc., New York).

Dado que la información suministrada por los proveedores del instrumento no era suficiente para ponerlo a punto y poder utilizarlo en su máximo potencial, antes de abordar los objetivos de la presente tesis doctoral, se contrastó con un instrumento de laboratorio, de alto coste y prestaciones científicas, que ya había sido utilizado en el trabajo de tesis doctoral de Riccioli C., (2012), realizado en el seno del mismo grupo de investigación (AGR-128).

El barrido espectral considerado de referencia para la evaluación de las medidas realizadas por el equipo no comercial, fue un equipo integral MatrixNIR (Malvern Instruments, Maryland, USA) que cubrió el rango NIR (900-1700 nm), equipado con un filtro de cristal líquido ajustable (LCTF) (Fig.8). Se usó con dos lámparas halógenas de 250 w en lugar de cuatro lámparas polarizadas (para igualar las condiciones de adquisición de imágenes con ambos equipos). La cámara y el filtro eran controlados por un ordenador integrado (Dell Inc., Round Rock, TX, USA. UU). La cámara dispone de

un sensor compuesto de Indio Galio, Arsenio (InGaAs), enfriado termoeléctricamente y capaz de adquirir imágenes de 256x320 píxeles y una relación focal de F/5.

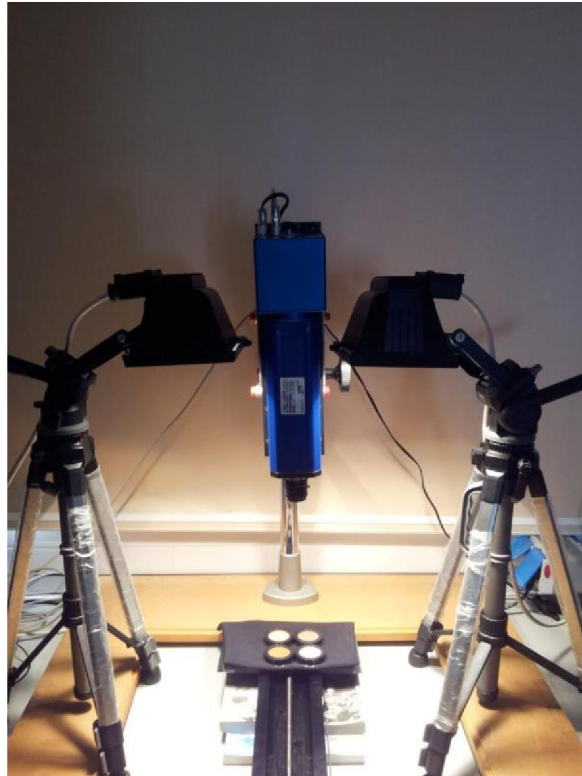


Figura 7. Equipo de barrido lineal modular no comercial.



Figura 8. Equipo Matrix NIR integral de laboratorio, de barrido espectral, con las dos lámparas halógenas de 250 w.

La resolución espacial fue establecida en $306 \mu\text{m}/\text{píxel}$ en ambos sistemas, variando la distancia al objetivo a la muestra y la velocidad de desplazamiento de la plataforma para el equipo de barrido lineal. Esta resolución era la mínima resolución permitida por el equipo de barrido espectral y la máxima del equipo de barrido lineal. Esto obligó a posicionar el portamuestras lo más lejos posible del objetivo en el equipo de barrido espectral, haciendo imposible el uso de las cuatro lámparas integradas del equipo comercial. La resolución espectral se fijó en $3,3 \text{ nm}$ en ambos sistemas (la preestablecida para el equipo de barrido lineal y la máxima para el equipo de barrido espectral). Aunque el rango espectral es el mismo para ambos equipos, se establecieron 70 longitudes de onda, tal y como recomiendan los fabricantes para el equipo de barrido espectral. Para ello, los valores que tomaron las 212 longitudes de onda en el equipo de barrido lineal se interpolaron linealmente para poder hacer comparables los tests. Se fijaron similares tiempos de exposición en ambos sistemas (128 ms para el equipo de barrido espectral y 110 ms para el equipo de barrido lineal). En estas condiciones, el tiempo de adquisición de las imágenes para el equipo de barrido espectral fue de 15 min y para el equipo de barrido lineal de 5 min, aproximadamente.

La saturación del detector fue del 80 % en ambos casos, permitiendo una suficiente intensidad evitando problemas saturación. Los rangos dinámicos eran de 12 bits, 0-4095, para el equipo de barrido espectral y 14 bits, 0-16383, para el equipo de barrido lineal. Todos los tests fueron realizados a 22 ± 1 ° C gracias al acondicionamiento de las salas donde se ubicaban ambos instrumentos.

Los píxeles anómalos se eliminaron mediante Matlab R2008a. Los píxeles con valores fuera del rango (NaN, Inf y -Inf) fueron sustituidos por el espectro promedio de cada imagen. También se consideraron píxeles anómalos aquellos con picos mayores que dos veces la desviación estándar de todos los valores.

Se efectuaron dos test de comparación. En primer lugar, se realizó una comparación de la nitidez espectral (“Sharpness of the spectra: Full Width at Half Maximum”). En segundo lugar, se compararon los niveles de la relación señal/ruido (Signal to Noise) entre ambos equipos.

5.1.1. Metodología.

Señal/Ruido (SNR).

El ratio señal/ruido (SNR) se define como la proporción existente entre la potencia de la señal que se transmite y la potencia del ruido que la distorsiona. El SNR cuantifica la bondad de un instrumento de medida. En este caso, se usó para caracterizar y comparar el comportamiento de ambos instrumentos. Sus valores se calcularon usando Matlab R2008a. El SNR se calculó aplicando la siguiente fórmula a los píxeles seleccionados de la manera que se explica más adelante:

$$SNR = \frac{\mu}{\sigma}$$

donde μ es la señal media para una longitud de onda dada y σ su desviación típica para una serie de mediciones.

Debido a que para el portamuestras del equipo de barrido espectral hay que tener en cuenta dos dimensiones y a su mayor tamaño, se tomaron cinco medidas de cada

cerámica de referencia (cerámicas de referencia con reflectancias estandarizadas de 5 %, 10 %, 20 %, 40 % y 60 %) en dos posiciones distintas del portamuestras: en total 10 medidas para cada cerámica. En la figura 9 se muestra una de las posiciones. Con el equipo de barrido lineal, debido a que su medida es unidimensional en el espacio, se tomaron diez medidas para cada cerámica de referencia en una sola posición sobre el portamuestras (plataforma del motor de pasos, en este caso), como puede apreciarse en la figura 10. Con cada medida y cambio de posición se tomaron unas nuevas referencias blanca y negra, para el cálculo de la reflectancia.

En ambos sistemas la región de interés (Region of Interest –ROI-) fue de 100 x 100 píxeles. A partir de este ROI, mediante interpolación entre los vecinos más cercanos (“nearest-neighbour interpolation”), se obtuvieron tres hipercubos más de cada hipercubo original, disminuyendo la resolución (2x2, 5x5 y 10x10 píxeles). De cada medida y resolución se seleccionaron aleatoriamente 10 píxeles mediante valores pseudoaleatorios en el equipo de barrido (“equipo de barrido pseudorandom values”) generados en Matlab R2008a. La señal y el ratio señal/ruido fueron calculados para cada resolución usando los 100 píxeles seleccionados del modo anteriormente descrito (100 píxeles originarios de 10 imágenes de la misma cerámica). En la Tabla 3 se hace un resumen de las combinaciones de disposiciones adoptadas.

Tabla 3. Resumen de las réplicas tomadas para cada uno de los equipos.

Resolución	Equipo de barrido lineal	Equipo de barrido espectral	
	(10 réplicas, 100 píxeles seleccionados)	(10 réplicas, 100 píxeles seleccionados)	
	(10 réplicas)	Posición c (5 réplicas)	Posición d (5 réplicas)
1x1	10 píxeles seleccionados / réplicas = 100 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles
2x2	10 píxeles seleccionados / réplicas = 100 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles
5x5	10 píxeles seleccionados / réplicas = 100 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles
10x10	10 píxeles seleccionados / réplicas = 100 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles	10 píxeles seleccionados / réplicas = 50 píxeles

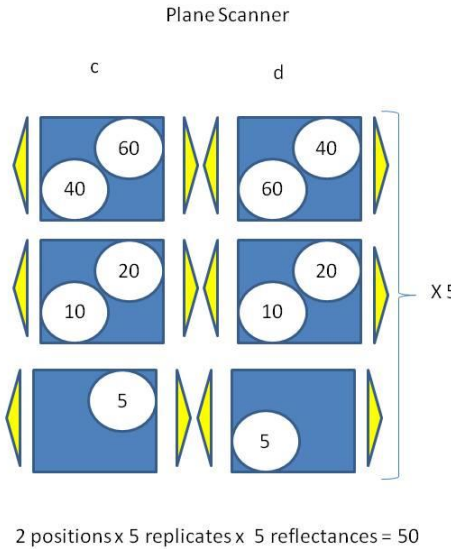


Figura 9. Posiciones de las cerámicas de referencia en el equipo de barrido espectral.

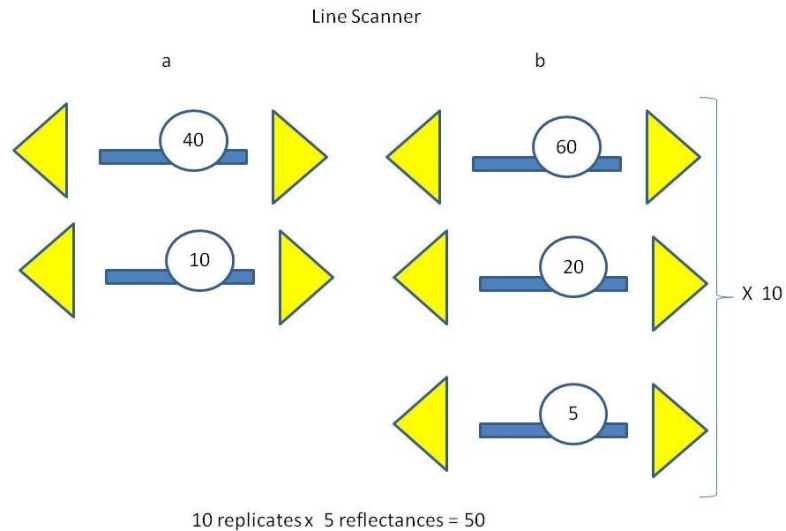


Figura 10. Posiciones de las cerámicas de referencia en el equipo de barrido lineal.

Nitidez spectral.

La “anchura a media altura”, abreviada FWHM (del inglés Full Width at Half Maximum) es una medida de la extensión de una función, que viene dada por la diferencia entre los dos valores extremos de la variable independiente en los que la variable dependiente es igual a la mitad de su valor máximo. Se usó este parámetro para comparar la nitidez de la señal midiendo una cerámica de referencia (“Oxide reference panel”).

5.1.2. Resultados de la comparación.

Señal/Ruido (SNR).

En las figuras 11 y 12 se representan las reflectancias de las cerámicas de referencia utilizadas para el cálculo del SNR (5 %, 10 %, 20 %, 40 % y 60 %). La reflectancia de cada cerámica de referencia se calculó con respecto a la cerámica de referencia de 99 % de reflectancia o blanco. Mientras que el blanco o referencia del 99 % en el equipo de barrido espectral es una imagen bidimensional de la cerámica, en el equipo de barrido lineal se usó un promedio de veinte líneas o “scans”.

Puede apreciarse cómo, pese a acercarse a su valor teórico de referencia, cuanto menor es la reflectancia, más se desvían con respecto a éste. En el equipo de barrido lineal, los valores obtenidos van disminuyendo con respecto al valor teórico a medida que disminuye la reflectancia de referencia; mientras que en el equipo de barrido espectral ocurre lo contrario. Pese a ser la cerámica de referencia del 99 % una cerámica estandarizada, su uso puede rebajar dicha reflectancia en alguna zona o zonas de la misma, por ejemplo en sus márgenes.

Este defecto se reflejaría más acusadamente en el equipo de barrido espectral debido a que su sistema de adquisición de imágenes recoge toda la superficie de la muestra. Esto explicaría la mayor reflectancia observada en las cerámicas de referencia para el equipo de barrido espectral. La ausencia de esta circunstancia, junto a una mayor sensibilidad del detector para intensidades de luz mayores, podría explicar el efecto contrario en el equipo de barrido lineal.

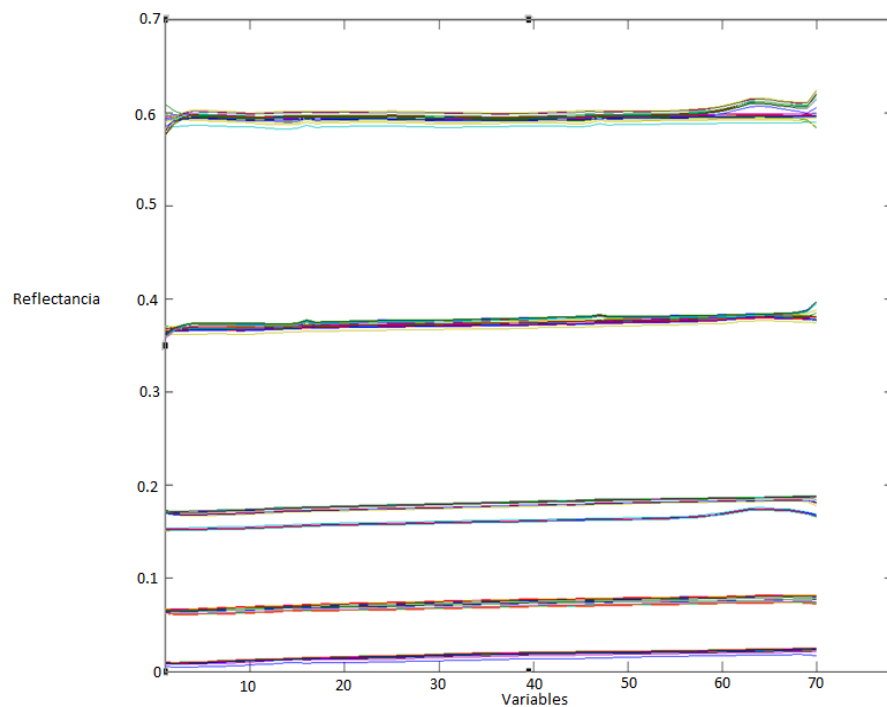


Figura 11. Señal con todos los píxeles (antes de la selección aleatoria de píxeles) para el equipo de barrido lineal.

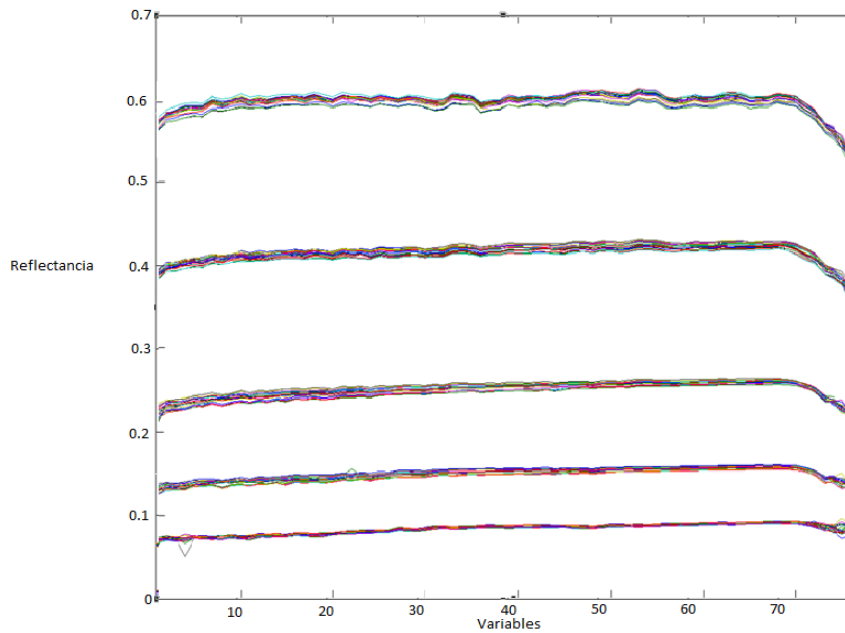


Figura 12. Señal con todos los píxeles (antes de la selección aleatoria de píxeles) para el equipo de barrido espectral.

En las figuras siguientes (de la 13 a la 20), se representan los valores de SNR para cada una de las 70 longitudes de onda, para cada resolución y nivel de reflectancia.

Se puede observar, en primer lugar, que cuanto menor es la resolución, mayor es el SNR. Esto se explica por la minimización del error al aumentar la población de píxeles promediados (mayor es la anulación de los errores de unos píxeles con respecto a otros).

En segundo lugar, cabe destacar que el ruido medido no es proporcional a la señal. En el equipo de barrido lineal hay un salto, mucho más acusado entre las referencias de 40 y 60; este fenómeno también se da en el equipo de barrido espectral, aunque el salto se produce entre la cerámica de 20 y la de 40. Esto nos indica la existencia de un ruido base independiente de la intensidad, probablemente derivado de la transformación de la señal de analógica a digital (A/D). Este ruido base llega a confundir las curvas SNR para las reflectancias de 5, 10 y 20 en el equipo de barrido espectral. Las últimas longitudes de onda deberían ser eliminadas cuando se trabaja con el equipo de barrido espectral ya que su SNR decae mucho más acusadamente que en el equipo de barrido lineal en el extremo superior de su espectro.

Por último, y en términos generales, el SNR del equipo de barrido lineal fue del orden de 1,7 veces el SNR del equipo de barrido espectral.

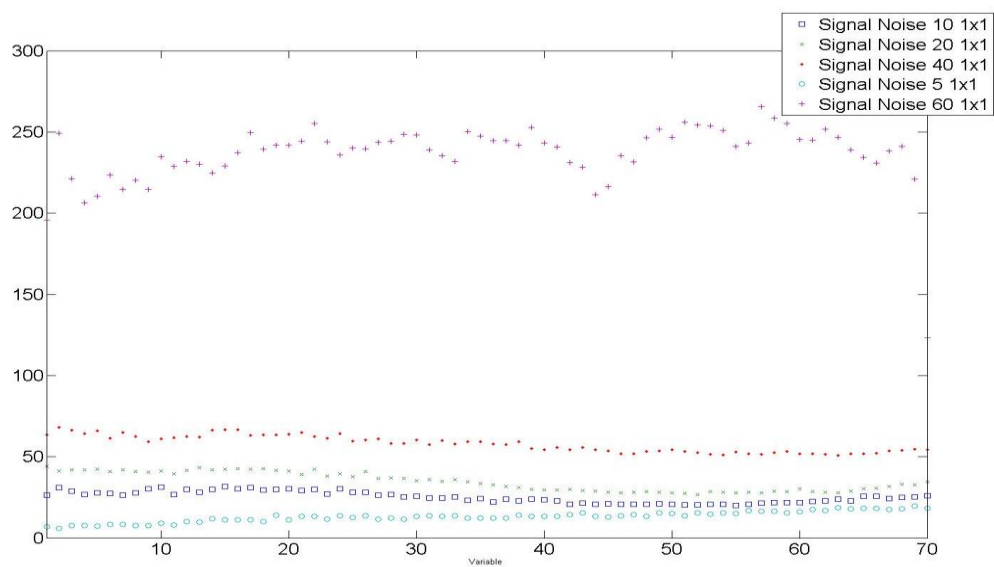


Figura 13. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la máxima resolución (1x1) en el sistema de barrido lineal.

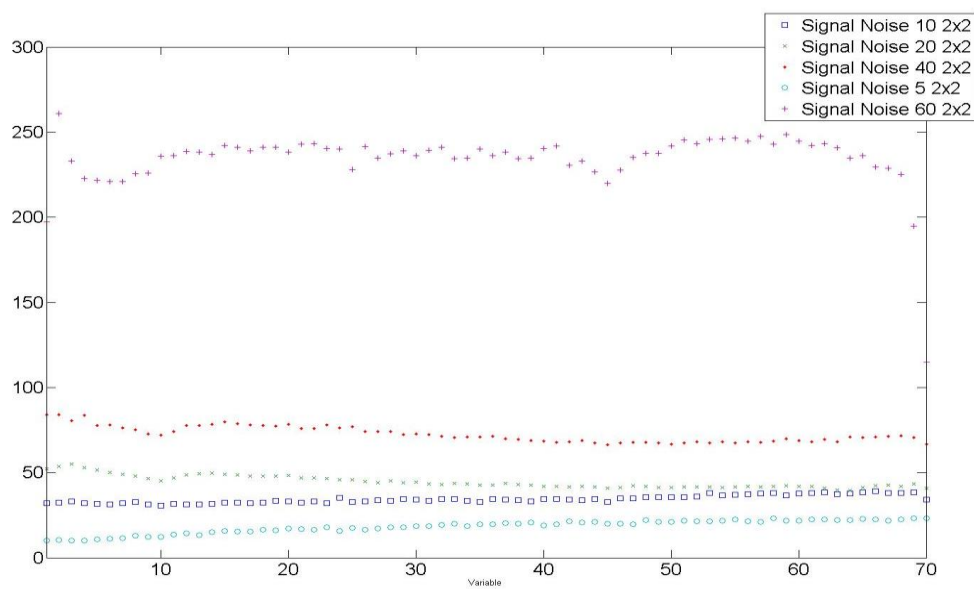


Figura 14. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 2x2 en el sistema de barrido lineal.

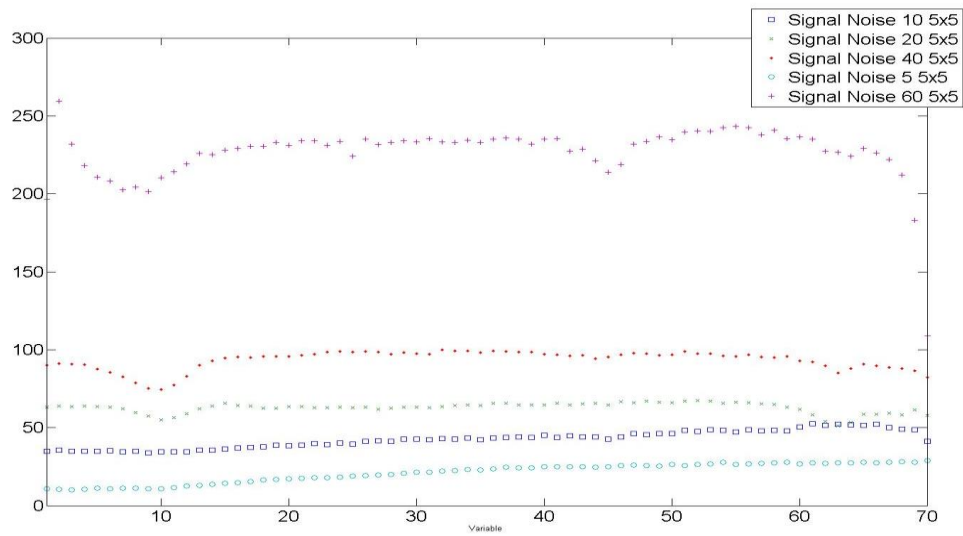


Figura 15. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 5x5 en el sistema de barrido lineal.

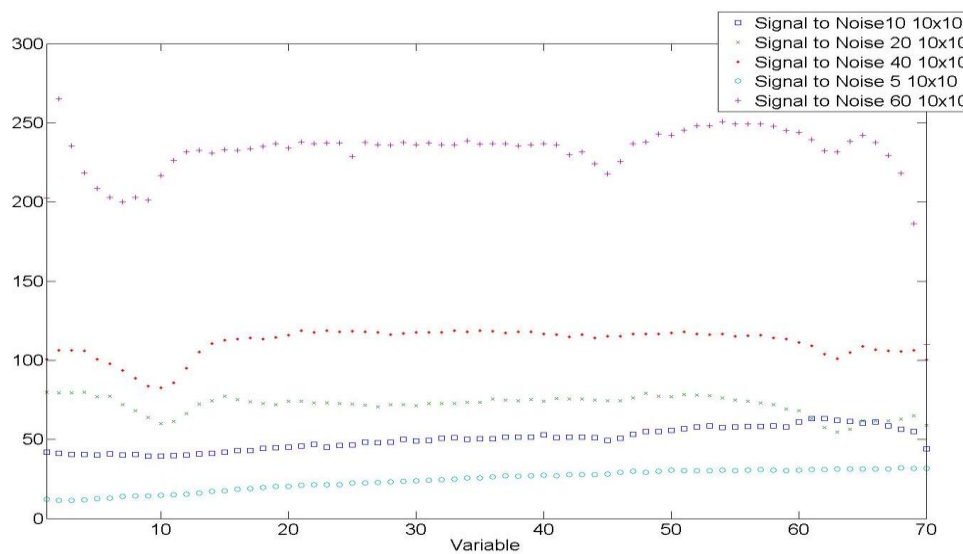


Figura 16. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 10x10 en el sistema de barrido lineal.

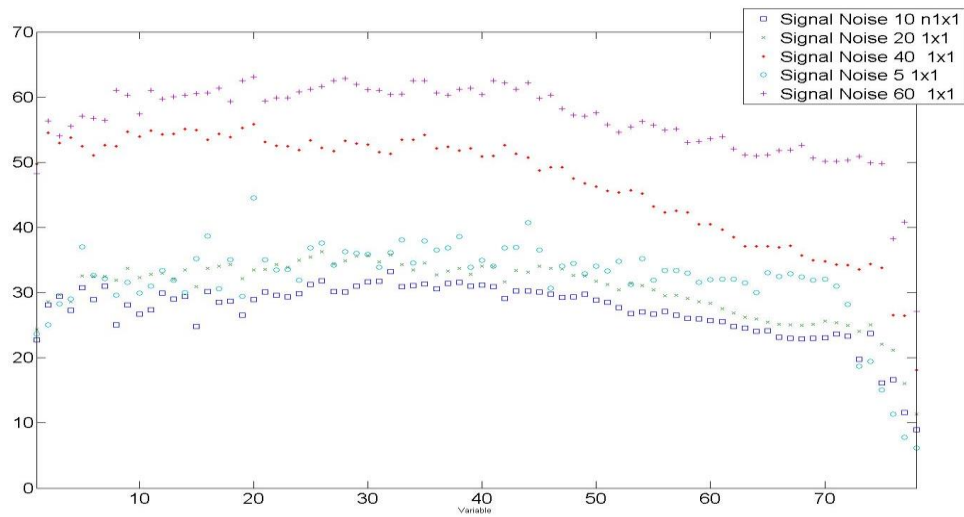


Figura 17. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la máxima resolución (1x1) en el sistema de barrido espectral.

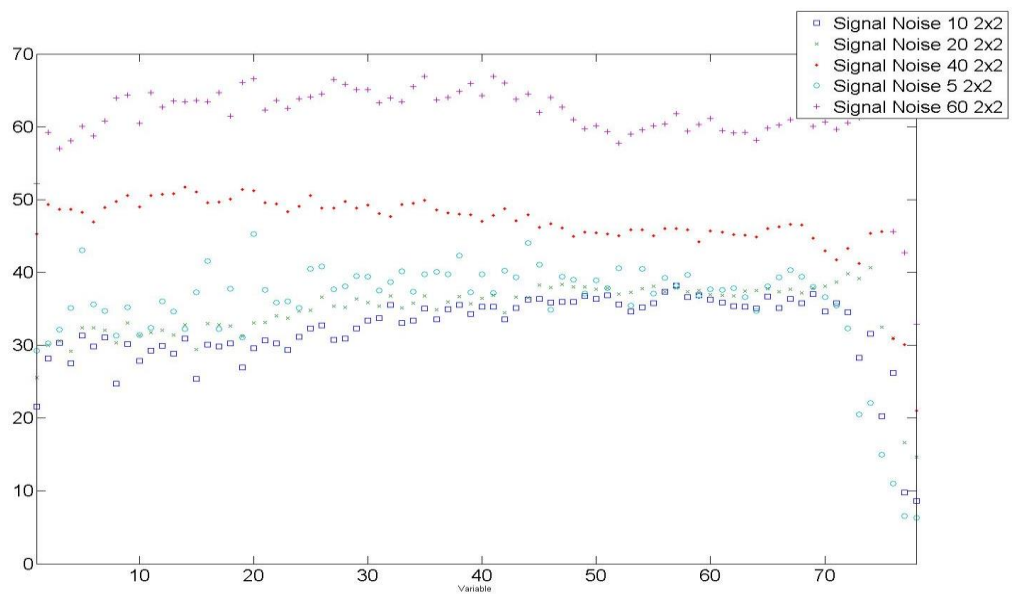


Figura 18. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 2x2 en el sistema de barrido espectral.

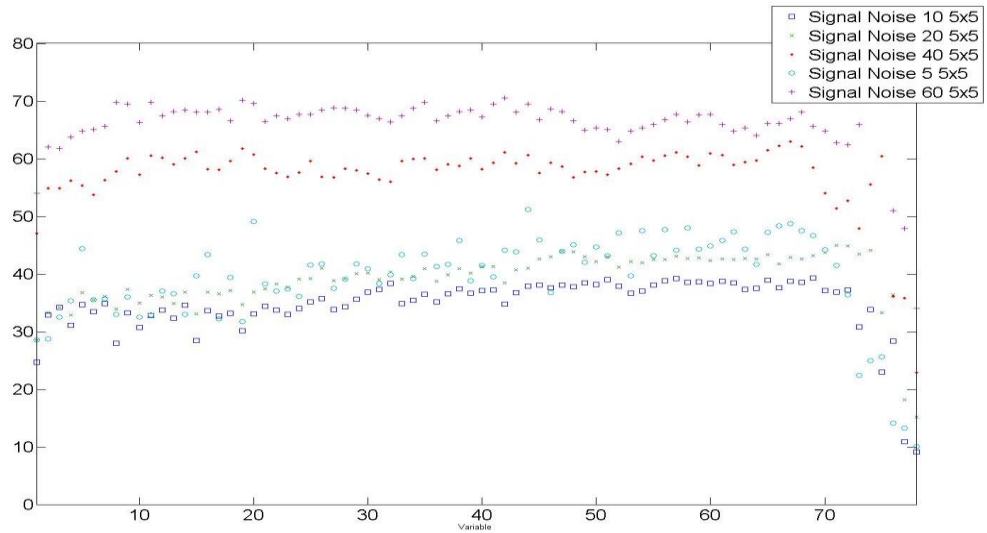


Figura 19. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 5x5 en el sistema de barrido lineal.

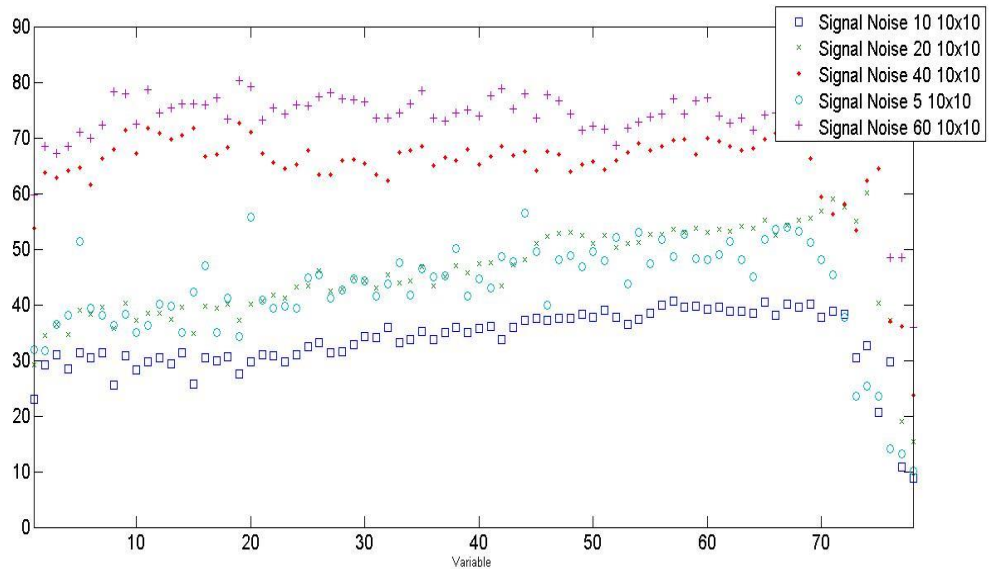


Figura 20. SNR para todas las longitudes de onda, y todas las cerámicas (5, 10, 20, 40, 60) a la resolución de 10x10 en el sistema de barrido lineal.

La FWHM se aplica a fenómenos tales como la duración de un pulso y la anchura espectral de fuentes utilizadas para comunicaciones ópticas y la resolución de espectrómetros. Cuanto menor es su valor, mayor es la nitidez de la señal del

instrumento. La FWHM del equipo de barrido lineal resultó inferior. Se obtuvieron los siguientes valores (Figura 16):

$$FWHM_{\text{EQUIPO DE BARRIDO ESPECTRAL}} = 1320,52 - 1259,23 = 61,29 \text{ nm}$$

$$FWHM_{\text{EQUIPO DE BARRIDO LINEAL}} = 1320,91 - 1260,03 = 51,88 \text{ nm}$$

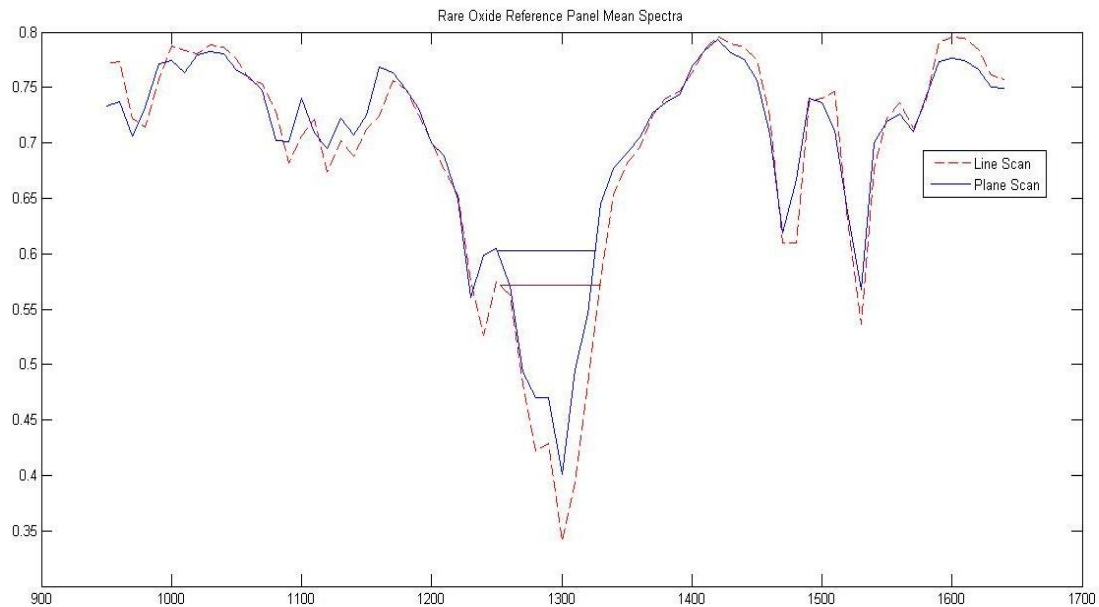


Figura 21. Nitidez espectral (Full Width at Half Maximum) para el equipo de barrido espectral y el de barrido lineal.

5.1.3. Conclusiones de la comparación.

- La adquisición simultánea de toda la superficie de la muestra en el equipo de barrido espectral da lugar a la adquisición de una reflectancia del 100 % o blanco menos uniforme que en el equipo de barrido lineal debido a irregularidades en la misma y a irregularidades en la iluminación. En el equipo de barrido lineal, esto se solventa gracias a la adquisición de solamente unas líneas (recordemos que su campo de visión es lineal). Ello hace que la irregularidad de la referencia del 100 % se minimice y la iluminación sea más homogénea cuanto menor sea el campo iluminado considerado.

- El ruido base es probable que provenga de la transformación de la señal de analógica a digital (A/D), lo que provoca que se confundan las curvas SNR para las reflectancias de 5, 10 y 20 en el equipo de barrido espectral, debido a la mayor proporción que supone el ruido con respecto a la señal en este equipo.
- Las últimas longitudes de onda deberían ser eliminadas (de la banda 70 en adelante) cuando se utiliza el equipo de barrido espectral, ya que su SNR decae mucho más acusadamente que en el equipo de barrido lineal en el extremo superior de su espectro.
- En términos generales, el SNR del equipo de barrido lineal fue del orden de 1,7 veces el SNR del equipo de barrido espectral; es decir, el ruido representa una menor parte de la señal en el equipo de barrido lineal.
- Aunque similares, la nitidez espectral del equipo de barrido lineal resultó mayor que la del equipo de barrido espectral.

En definitiva, el equipo de barrido lineal no comercial, no sólo se mostró un equipo comparable en su funcionamiento al equipo de barrido espectral comercial de laboratorio, sino que resultó ser superior en su comportamiento para los parámetros analizados.

Asimismo, las imágenes obtenidas con el equipo de barrido lineal se guardan en formatos de tres dimensiones de bandas intercaladas por líneas (“band interleaved by line” –BIL-). Este formato almacena los datos de todas las bandas desde el primer escaneo, seguido por todas las bandas del segundo escaneo. Los datos de los siguientes escaneos son intercalados de la misma forma. Un espectro completo puede dibujarse en cada píxel, e imágenes de dos dimensiones espaciales pueden ser compiladas de todos los escaneos a una longitud de onda determinada. Por otro lado, los datos del barrido lineal resuelto espacialmente suelen almacenarse en una imagen de dispersión de dos dimensiones, de la que se pueden extraer los espectros a diferentes distancias del detector y los perfiles espaciales a distintas longitudes de onda. Este tipo de formato es mucho más manejable que el formato obtenido mediante el equipo comercial, que requiere una transformación previa con un programa específico.

5.2. “Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288





Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras

Cristóbal Garrido-Novell^{a,*}, Dolores Pérez-Marin^a, Jose M. Amigo^b, Juan Fernández-Novales^a, Jose E. Guerrero^a, Ana Garrido-Varo^a

^a Faculty of Agriculture and Forestry Engineering, Non-Destructive Spectral Sensors Unit, University of Cordoba, Spain

^b Faculty of Life Sciences, Department of Food Science, Quality & Technology, University of Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 27 September 2011

Received in revised form 19 April 2012

Accepted 27 May 2012

Available online 7 June 2012

Keywords:

Hyperspectral imaging

Imaging processing

RGB imaging

Apples

Color evolution

Shelf-life

Linear discriminant analysis

PLS-DA

ABSTRACT

The potential of RGB digital imaging and hyperspectral imaging (900–1700 nm) was evaluated for discriminating maturity level in apples under different storage conditions along the shelf-life. Segmentation, preprocessing and partial least squares-discriminant analysis (PLS-DA) were used for hyperspectral data analysis, while illumination correction, dimensionality reduction and linear discriminant analysis (LDA) were used for RGB data analysis. The results showed that hyperspectral discrimination classified different storage regimes better than RGB, with an overall success rate of 95.83%. In addition, color evolution of apples during shelf-life under different storage regimes was modeled using RGB zero and first order regression models, fitting better to a first order kinetic model.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

A healthy, fresh and colorful look is the first pleasurable sensation that consumers perceive in fruits, affecting the rest of the sensations perceived. For this reason, it is essential to quantify this first sensation. The color constitutes the first relationship established between the product and the consumer, being essential for making purchase decision.

In the ripening process, the degradation of surface chlorophyll of the fruit takes place while colored pigments, both carotenoids (yellow) and anthocyanins (red), are being synthesized. That is, two colors may appear: the first one, yellow, is the background color; and the second one, red, is the surface color. The color of Golden Delicious apples (*Malus domestica*) varies from green to yellow-green tones showing a slight pinkish blush.

Computer vision is a rapid, economic, consistent and objective inspection technique, which has expanded into many diverse industries. Computer vision has been used for such tasks as shape classification, defects detection, and quality grading and variety classification. Color evolution under different thermal treatment in apples has been evaluated in dry, sliced and squeezed apples and it has only been modeled in Champion and Jonajold fresh

intact apples during shelf-life (Garza et al., 1996; Krokida et al., 1998; Dobrzanski and Rybczynski, 2002), using all of them L*a*b* color space. Dobrzanski et al. (2006) showed that estimation of fruit quality based on L*a*b* system describing coordinates of color could be useful in connection with marketing, for monitoring consumer preferences and assessing the products after storage and at shelf-life. Dobrzanski and Rybczynski (2002) concluded that determination of fruit quality based on 3-d coordinates of color should be improved to clearly define blush and ground color of apple skin.

Hyperspectral imaging provides physical and chemical characteristics of an object, like NIRS-technology, as well spatial information. In recent years, hyperspectral imaging techniques have been implemented in various applications in the horticultural sector. There are numerous papers on fruits that evaluate the ability of imaging and spectral information for apple grading according to different defects. However, grading according to appearance evolution under different storage conditions has only been studied using colorimeters, not RGB imaging, nor hyperspectral, as far as we know. In addition, given the different experimental conditions between each trial is difficult to draw conclusions and recommendations on the advantages of using spectral imaging sensors in relation to the conventional image for apples.

Before taking into account any result of the study, it can be assessed that the much greater amount of information and the ability to quantify other attributes, like acidity, sugar content,

* Corresponding author.

E-mail address: crisgano8459@hotmail.com (C. Garrido-Novell).

etc., (Lu, 2003; Mehl et al., 2004; Ariana et al., 2006; Peng and Lu, 2008; ElMasry et al., 2009) of hyperspectral imaging system is unreachable by RGB imaging system, especially in the NIR range. However, applications using hyperspectral imaging system in the food industry for inspection of agricultural products are still very limited due to its high time requirement for the acquisition, processing and classification of images, and high prices (Mehl et al., 2004; Kim et al., 2008). Nowadays, however, cheaper and faster hyperspectral equipment is being released, starting to be economically interesting its implementation. For this reason developing new algorithms in order to process faster the information contained into a hypercube is one of the main goals of this area.

Therefore, the goal of this paper were to evaluate and modeling color evolution under different storage regimes, simulating different market storage conditions with RGB and hyperspectral imaging, comparing its ability to sort apples according to color and storage conditions using the same collection of samples.

2. Materials and methods

2.1. Samples

Apples, cultivars 'Golden Delicious', obtained in a retail store (Mercadona Company, Córdoba, Spain). The samples were transported to the University of Cordoba (UCO) facilities and stored in a cold room (3 °C). Apples, even individual apples, displayed color variations that may have resulted from differential environmental growth conditions causing variations in pigmentation and ripeness. A total of 40 apples without any physical defects (that were visually identifiable) were randomly selected for imaging study. Hence, they constituted a range of the natural commercial color variations.

2.2. Experimental design

The experimental design was done following the isochronous method (Lamberty et al., 1998; Amigo et al., 2004). The whole apples were divided into 8 five-apple groups. All of them were stored into a cold room (3 °C and 80% of RH). These conditions were chosen as reference. All changes in color will be related to the color at this stage. Every three days a group was put into normal conditions in a temperature controlled room (20 °C and 40% HR), so in two week – commercial apples shelf-life in supermarkets – all the apples were out of the cold room. The first group was at room temperature conditions since the first day and another group was saved under freezing conditions (−18 °C) (Table 1). This design try to emulate all different storage conditions, that is, under cold and humidity controlled conditions and room conditions, passing through the intermediate ones. Isochronous design has the advantage over other experimental design used in previous studies on evolution of quality under storing conditions that allows making all the measurements at the same time, avoiding laboratory and measurement errors.

2.3. Apparatus

The following instrumentation was used:

- RGB system: a Nikon digital camera (Coolpix L3, 6.2 effective Megapixels CCD); green and red references; a 60 W halogen illumination.
- Hyperspectral imaging system: spectrophotometer Specim (ImSpector V10E) in the NIR range (900–1700 nm) with a spectral resolution of 5 nm, push broom XEVA –FPA-1.7–320–XeniCs CCD camera (spatial resolution = 320 × 256) and a exposition time of 1 μmin–5 min, acquisition system via USB, and halogen illumination with two 250 w lamps.

2.4. Experimental

2.4.1. RGB images

Scanning of the whole set of apples was carried out on three sides of the apples after being cleaned with paper towel and the result was treated as a bitmap image. An amount of 2592 × 1994 pixels per photo per apple face was acquired. All the data points in a region of interest (ROI) can be represented by the mean and standard deviation of the color intensity values:

$$mean_{Image} = MCV = \frac{\sum_{x=1}^X \sum_{y=1}^Y p_{xy}}{X + Y} \quad (1)$$

$$s_{Image} = \sqrt{\frac{\sum_{x=1}^X \sum_{y=1}^Y (p_{xy} - mean_{Image})^2}{X + Y - 1}} \quad (2)$$

where X and Y are the number of pixels in the horizontal and vertical axes respectively, p_{xy} denotes the pixel p in the x th – y th position of the ROI, $mean_{Image}$ is the mean color value (MCV) and s_{Image} the standard deviation of the histogram.

2.4.2. Stability of light correction

There is a need to ensure that changes in the apple image come only from the ripening process not from the illumination of the associated measurement process. In this case a color standard with 2 colors (red and green) was used (Fig. 1). This provides information to assess if there is any change in lighting conditions (Vidal et al., 2011a,b).

To make a correction it should be considered that the original images of the color standard (I_1 , I_2 , etc.) are formed by m pixels, but once they are transformed to the RGB model and unfolded, they form a matrix of m columns (the number of pixels) and 3 files (the RGB colors); that is, any of those matrices can be represented by $I_{(m,3)}$.

Bold characters represent matrices and superscript T means the transpose. The dimensions of matrices are given as a subscript. To

Table 1
Storage design for isochronous measurements.

	0 group	1st group	2nd	3rd group	4th group	5th group	6th group	Frozen apples
1st three days	20 °C ^a , 40% ^b	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	−18 °C
2nd three days	20 °C, 40%	20 °C, 40%	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	−18 °C
3rd three days	20 °C, 40%	20 °C, 40%	20 °C, 40%	3 °C, 80%	3 °C, 80%	3 °C, 80%	3 °C, 80%	−18 °C
4th three days	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	3 °C, 80%	3 °C, 80%	3 °C, 80%	−18 °C
6th three days	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	3 °C, 80%	3 °C, 80%	−18 °C
7th three days	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	20 °C, 40%	3 °C 80%	−18 °C

^a Room temperature.

^b Relative humidity.



Fig. 1. Image of a sample. In the background both red and green references can be seen. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

correct $I_{2(m,3)}$ into $I_{1(m,3)}$, when necessary, the matrix $M_{(3,3)}$ must be obtained:

$$I_{2(m,3)} \cdot M_{(3,3)} = I_{1(m,3)} \quad (3)$$

The correction can be expected to conform to a linear model and then:

$$M_{(3,3)} = [I_{1(3,m)}^T \cdot I_{1(m,3)}]^{-1} I_{1(3,m)}^T \cdot I_{2(m,3)} \quad (4)$$

Once $M_{(3,3)}$ is known, it can be applied to correct the image of the corresponding apple image, regardless the number of pixels. If the image, transformed to the RGB model and unfolded, is represented by $P_{2(n,3)}$ (n pixels), the corrected image will be represented by:

$$P_{1(n,3)} = P_{2(n,3)} \cdot M_{(3,3)} \quad (5)$$

The image $P_{1(n,3)}$ can be compared with any other apple image or corrected according to Eq. (3). Obviously, the matrix $M_{(3,3)}$ in Eq. (4) changes for each plate, whenever $I_{2(m,3)}$ differs significantly from $I_{1(m,3)}$ owing to changes in lighting conditions, but if they are stable, $M_{(3,3)}$ is the unity matrix.

2.4.3. Hyperspectral images

2.4.3.1. Acquisition. Two hyperspectral images, from different sides after been cleaned with paper towel, were taken from each apple. The image acquisition was carried out at room temperature. Each image was acquired in the spectral range of 900–1700 nm with 3.15 nm intervals between contiguous bands with a total 208 bands. To correct the acquired image a white and dark reference images were captured. The dark reference image of 0% reflectance was obtained by turning off the light source and completely closing the lens of the camera, while the white reference image of 100% reflectance was acquired for a white Teflon calibration board. A black sample holder was designed and used to provide contrast between apples and the background.

2.4.3.2. Preprocessing. Segmentation consisted in separating the apple body from the background, the brightness and the peduncle, according to its spectra (Fig. 2). First, a subtraction ratio was made between the image taken at 1065 nm and the image taken at 1578 nm multiplied by two. These wavelengths were selected because in this two points of the spectrum the difference in reflectance between the apple and both the background and the peduncle were maximum (Fig. 3). The subtracting term was

multiplied by two in order to decrease the reflectance value of the bright zone under the value obtained for the rest of the apple (Fig. 4) (ElMasry et al., 2011). Once the reflectance values of the undesirable zones had been reduced considerably from the values of the ROI, Otsu's algorithm (Otsu, 1979) was applied instead of applying a manual binary threshold in order to make the process as automatic as possible.

Although brightness could have been avoided by decreasing the illumination intensity, the nearest zone of the apple to the lens was saturated in order to decrease shadows in lower zones due to the spherical effect. In this way, although the highest zone of the apple was not measured, the total zone measured was increased by including darker zones that in other case would have been confounded with the background.

Otsu's method searches for the threshold, that is the value of reflectance that divide the whole pixel into two group or segments, that minimizes the intra-segment variance, at the same time that maximizes the inter-segment variance. This is carried out by maximizing the ratio between the inter-segment variance and intra-segment variance.

Once the binary image, resulted from the Otsu algorithm, has been obtained, all the images, one for each wavelength, are masked by it. Applying the mask, all the pixels with one value are maintained, while the zero ones are removed from the hypercube. The spectra were obtained by averaging all the pixels maintained after masking (ElMasry et al., 2009).

Acquisition and preprocessing of data were realized with two programs: MATLAB (Version R2008a) and a program based in a SDK (Software Development Kit) provided by the camera manufacturer on a Microsoft MS Visual Basic (Version 6.0) platform in the MS Windows operating system.

2.5. RGB discriminant analysis

Linear discriminant analysis (LDA) was used for classifying RGB data of the apple images by the groups indicated in Table 1. Leave-one-out cross validation method was considered to validate the classification model. Each time, one sample was taken out and LDA was established for the remaining samples, and the model was then used to predict the sample left out. Thereafter, the sample was placed back into the model, and another sample was left out. This procedure was repeated until all samples had been left out once. LDA derives linear combinations of the independent variables that discriminant between a priori defined groups in such a way that maximum discrimination is achieved. This is accomplished by maximizing the ratio of the between-class distance against within-class distance. It was implemented using Minitab (release 16, Minitab Inc.).

2.6. Hyperspectral discriminant analysis

Discriminant models were constructed to classify apples by the groups indicated in Table 1, using PLS-DA for supervised classification, applied to the average spectra resulted from the segmentation process. Specifically, the PLS2 algorithm was applied, using the 'Discriminant Equations' option in the WINISI version 1.50 software package.

All models were developed using five cross-validation groups (i.e. the calibration set is partitioned into five groups; each group is then predicted using a calibration developed on the other samples) and a spectral range from 900 to 1700 nm, eliminating signal noise at the beginning and end of the spectral range. In all cases, the standard normal variate and detrending methods were applied for scatter correction (Barnes et al., 1989) and four derivative mathematical treatments were tested: 1,5,5,1; 2,5,5,1; 1,10,5,1; 2,10,5,1.

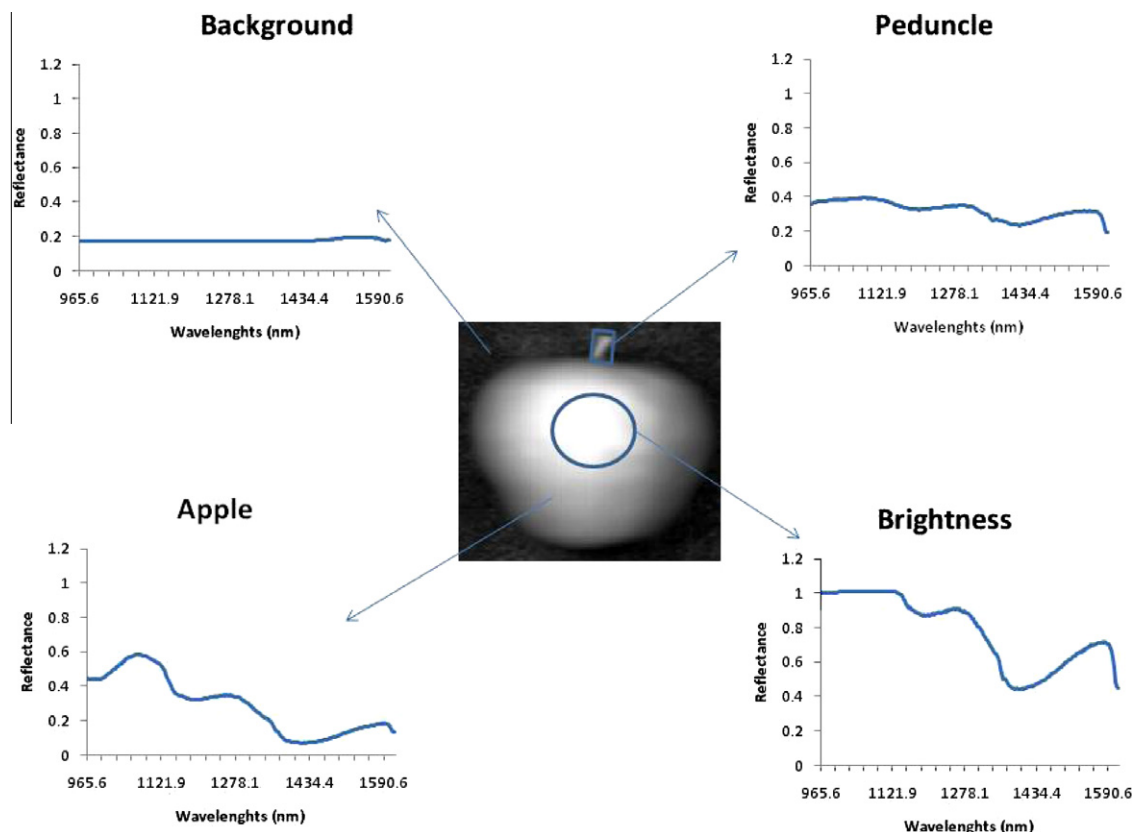


Fig. 2. Spectra of each different segment of the apple image.

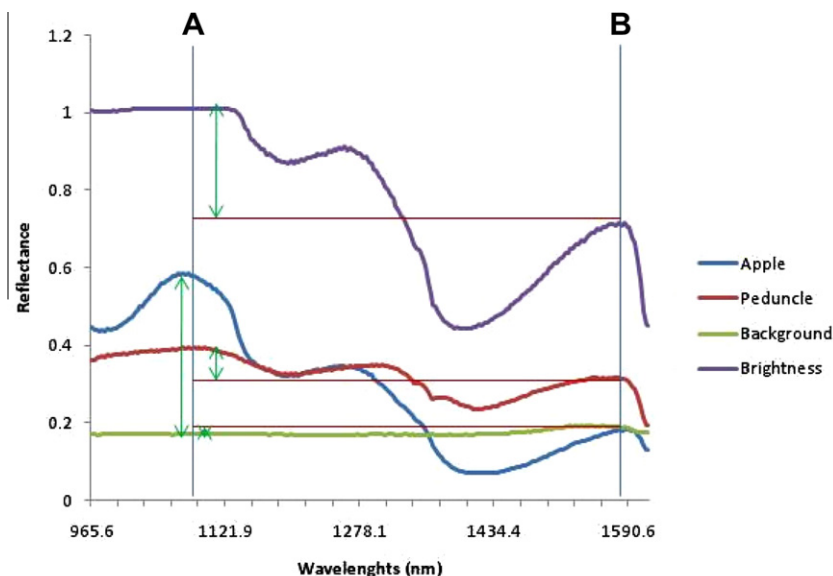


Fig. 3. Differences between the reflectance values for each segment spectra at 1065 and 1578 nm.

3. Results and discussion

The average intensity values of each sample obtained from the three channels compounds of the RGB unfolded image are presented in 2-d view versus the group (Fig. 5). Color evolution, which cannot be appreciated at first glance, is represented, in such a short time (15 days) by a clear tendency in the average red and green values from the decomposed RGB images (Fig. 5). However, the blue channel did not provide extra

significant information and differences between apples were not relevant. In fact, [Sudhakara Rao et al. \(2004\)](#) experimentally found for several varieties that the apple skin has negligible blue color component in the RGB color space. It can be appreciated that the cold storage during the shelf-life of apples stops the color evolution while the room condition storage accelerates the evolution (Fig. 5).

The values of red were in the range from 89 to 168, being the higher values from the apples that spent more time under room

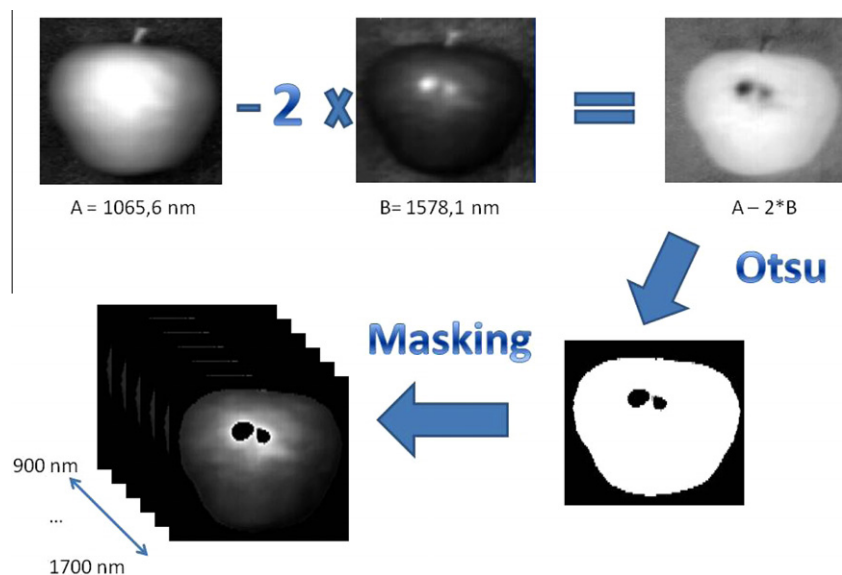


Fig. 4. Segmentation procedure.

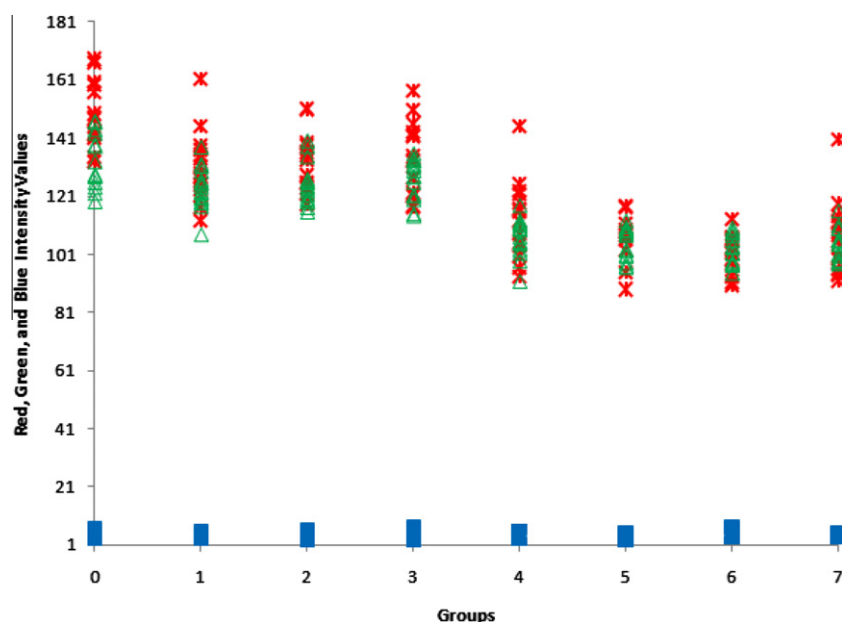


Fig. 5. Evolution of red, green and blue channel according to group number. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

conditions and the lower ones from the cold stored apples. Similarly, the green value range went from 91 to 146, increasing with the out of cold storage time. It can be appreciated that while neither red or green are predominant in the apples stored under cold condition, red is getting more and more predominant with the increasing time under room condition storage, although both red and green color present the same tendency. This can be seen in the slopes, and the ordinate origin of regression models obtained (Table 2). However, the blue component with a 7–2.5 range, is the least important channel among the three one throughout the process.

The two kinetics models tried to evaluate color evolution were zero and first order models, expressed by the following equations:

- Zero order ($dC/dt = k_0$): $C = C_0 + k_0 \times t$, rate of color changing independent of any factor.

- First order: ($dC/dt = k_1 \times C$): $C = C_0 \times \exp(k_1 \times t)$, rate of color changing dependent on t .

where C is the value of the color, in our case in RGB scale and C_0 is the initial color and t is the thermal treatment (time in cold storage) and k is the constant of the reaction velocity. The parameters for the linear regression are shown in Table 2.

The best regression results were obtained with the first order kinetic model, showing a 15% of the range standard error and a R^2 that explains 63% of the variance, for the red and green channel. However, the regression for blue channel shows there is no relation between the blue color and the evolution of the apples color under different thermal treatment.

Dobrzanski and Rybczynski, (2002) obtained a high slope (0.65) and determination coefficient ($R^2 = 0.56$) for chromaticity coordinate (b^*) – position between blue and yellow – linear regression

Table 2
Regression models for color evolution.

	Model	R^2	Standard error	Range
Red	$C = C_0 + k_0 \times t$	0.59	12.25	89–168
	$C = C_0 \times \exp(k_1 \times t)$	0.63	12.13	
Green	$C = C_0 + k_0 \times t$	0.63	8.2	91–146
	$C = C_0 \times \exp(k_1 \times t)$	0.63	8.17	
Blue	$C = C_0 + k_0 \times t$	0.006	2.3	2–7
	$C = C_0 \times \exp(k_1 \times t)$	0.01	0.88	

Table 3
Classification results from linear discriminant analysis of the different groups using RGB (RG) imaging.

B \ A	0 group	2nd group	4th group	6th group	Frozen apples
0 group	10	3	0	0	0
2nd group	5	11	1	0	0
4th group	0	1	6	6	0
6th group	0	0	6	9	2
Frozen apples	0	0	1	0	13

Percentage of correctly classified of the model after cross validation: 66.2%.
Qualitative groups (regime of storage): A, actual group; B, predicted group.

Table 4
Percentage of apples correctly classified using data from hyperspectral imaging (PLS-DA).

B \ A	0 group	2nd group	4th group	6th group	Frozen apples
0 group	7	0	0	0	0
2nd group	0	9	0	0	0
4th group	2	0	10	0	0
6th group	0	0	0	10	0
Frozen apples	0	0	0	0	10

Percentage of correctly classified of the model after cross validation: 96%.
Qualitative groups (regime of storage): A, actual group; B, predicted group.

according to L*a*b* system. Although they tried exponential regression too, all changes of color during display were well described by liner regression, while the exponential model indicated more closely the influence of time after bruising, point that is not treated in this paper, on all color coordinates. Similarly, RGB regression models carried out show not significant difference between linear and exponential regression, although the last one is slightly better (Table 2). On the other hand, while the better results obtained by Dobrzanski and Rybczynski, (2002) are for chromaticity coordinate (b^*), the best results for RGB color system are in the green channel.

3.1. Grading of apples according to color

Before carrying out any RGB or Spectra analysis the number of different groups of the experiment were evaluated by ANOVA, an analysis of whether or not the means of defined groups are all equal, combined with a Tukey test, which provide classification information for the defined group design with confidence intervals for differences between means of pairs of groups. The ANOVA and Tukey analysis revealed that there were four groups. For the grouping analysis and the following classification modeling the freeze apples group, which was excluded from the color evolution model, is introduced for its interest.

LDA is a tool that provides automatic object classification (Mignani et al., 2005). It finds optimum boundaries among classes by maximizing the ratio of between-class variance and minimizing the ratio of within-class variance. Red and Green values from RGB images were used as input to LDA, instead of the three color values, due to the irrelevant information given by blue channel. Table 3 shows the classification results in the form of a confusion matrix (Table 3).

The numbers of correctly classified groups are shown on the diagonal, whereas off-diagonal numbers denote misclassifications. Percentage of correctly classified samples of the model was 66.2%, while percentage of correctly classified sample by group was higher than 60% in all the representative groups, except for the 4th. The inability to correctly classify the fourth group is due to the confusion derived from the position of that group, in the center of the sample set.

3.2. Classification of samples from spectral data

In order to compare RGB classification model with hyperspectral classification model, the number of groups assumed in that model is assumed in this one too. Spectral variations were correlated with those five categories established. The results obtained with the best classification model, using PLS-DA, for different types of storage, using various math treatments are shown in Table 4.

It is seen that PLS-DA model developed for group recognition was able to classify all the groups with overall classification accuracy of 100 except for 0 group, which showed a 78% classification rate, confirming that the spectral range of the camera contain information enabling apples to be classified by regime of storage. A total of 95.83% of samples were correctly classified.

The results suggest that hyperspectral imaging has the potential to discriminate different storage condition in apples. To prove the relevance of using hyperspectral imaging system over other systems, classification results can be compared in Tables 3 and 4. While overall accuracy in RGB classification reached to 66.2%, it almost reached 100% in hyperspectral case. Moreover, while the worst classified group had a 43% correctly classified samples, a 78% correctly classified samples were managed in the hyperspectral case.

Unay et al. (2011) introduced an automatic grading for bi-colored apples by multispectral (centered at 450, 500, 750 and 800 nm with respective bandwidths of 80, 40, 80 and 50 nm)

machine vision. In the proposed solution they, firstly, grade into two categories: defective and healthy, obtaining 93.5% accuracy, using LDA. However, for a multi-category LDA classification, distinguishing 4 qualities, 83.3% accuracy is managed, with 50.9% accuracy for the worst group, while the whole spectra from 900 to 1700 nm provide us, using a PLS-DA, 95.83% accuracy for five groups and a 78% accuracy for the worst group. In the same research the accuracy rate was improved up to 85.6% accuracy, with 65.5% accuracy for the worst group, by designing a cascade multi-classifier group, while direct PLS-DA hyperspectral classification keeps better.

4. Conclusion

RGB models developed showed firstly that color evolution during market shelf-life can be quantified in apples with RGB digital cameras; and secondly that color evolution can be explained by a zero order kinetic model. However, the results demonstrated the greater potential of hyperspectral imaging compared with conventional RGB imaging, including in a supposedly adverse wavelength range – although the main changes could be attributed to visible wavelength range changes, the hyperspectral camera with a 900–1700 nm wavelength range showed much better discrimination rates.

The qualitative power of the technique joined to the possible quantitative analysis will have to be considered in the future. The conclusions reached by this study promote further research in order to evaluate discriminative potential in case of bigger and more variable set, that is, different varieties (two color and one color apples), longer storage periods, and on-tree conditions, with even less stable light condition, including other quantitative parameters.

Acknowledgements

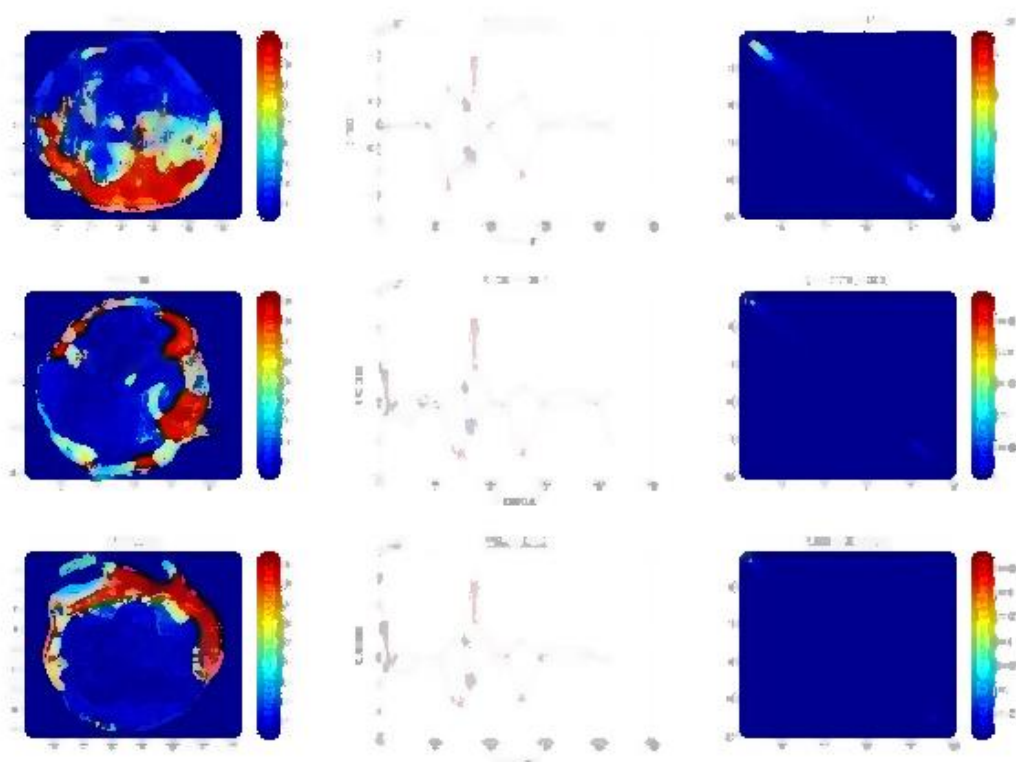
This research was carried out within the framework of the Research Excellence Program of the Andalusian Regional Government (Project No. P09-AGR-5129). The authors would like to thank Dr. Moon Kim (USDA-ARS, Beltsville) for his invaluable assistance in the setting and fine tuning of the hyperspectral imaging system.

References

- Amigo, J.M., Arana, G., Etxebarria, N., Fernández, L.A., 2004. Emerging needs for sustained laboratory reference materials. *Trends in Analytical Chemistry* 23 (1), 80–85.
- Ariana, D.P., Lu, R., Guyer, D.E., 2006. Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture* 53 (1), 60–70.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* 43 (5), 772–777.
- Dobrzanski, B., Rybczynski, R., 2002. Colour change of apple as a result of storage, shelf-life, and bruising. *International Agrophysics* 16, 261–268.
- Dobrzanski, B., Rabcewicz, J., Rybczynski, R., 2006. Handling of Apple transport techniques and efficiency vibration, damage and bruising texture, firmness and quality. *Institute of Agrophysics of Polish Academy of Sciences*. Available at <http://www.ipan.lublin.pl/mat_coe/mat_coe27.pdf>. Accessed on 30.03.2011.
- ElMasry, G., Wang, N., Vigneault, C., 2009. Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neuronal networks. *Postharvest Biology and Technology* 52 (1), 1–8.
- ElMasry, G., Iqbal, A., Sun, D.W., Allen, P., Ward, P., 2011. Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food Engineering* 103 (3), 333–344.
- Garza, S., Giner, J., Martín, O., Costa, E., Ibarz, A., 1996. Colour, sugars and HMF evolution during thermal treatment of apple juice. *Food Science and Technology International* 2 (2), 101–110.
- Kim, M.S., Lee, K., Chao, K., Lefcourt, A.M., Jun, W., Chan, D.E., 2008. Multispectral line scan imaging system for simultaneous fluorescence and reflectance measurements of apples: multitask apple inspection system. *Sensing and Instrumentation for Food Quality and Safety* 2 (2), 123–129.

- Krokida, M.K., Tsamia, E., Maroulis, Z.B., 1998. Kinetics on color changes during drying of some fruits and vegetables. *Drying Technology. An International Journal* 16 (3–5), 667–685.
- Lamberty, H., Schimmel, Pauwels, J., 1998. The study of the stability of reference materials by isochronous measurements. *Fresenius Journal of Analytical Chemistry* 360 (3–4), 359–361.
- Lu, R., 2003. Detection of bruises on apples using near-infrared hyperspectral imaging. *Transactions of the American Society of Agricultural Engineers* 46 (2), 523–530.
- Mehl, P.M., Chen, Y.R., Kim, M.S., Chan, D.E., 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering* 61 (1), 67–81.
- Mignani, A.G., Ciaccheri, L., Cimato, A., Attilio, C., Smith, P.R., 2005. Spectral nephelometry for the geographic classification of Italian extra virgin olive oils. *Sensors and Actuators B: Chemical* 111–112, 363–369.
- Otsu, N., 1979. A threshold selection method from gray-level histogram. *IEEE Transactions on System Man Cybernetics* 9 (1), 62–66.
- Peng, Y., Lu, R., 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biology and Technology* 48 (1), 52–62.
- Sudhakara Rao, P., Gopal, A., Revathy, R., Meenakshi, K., 2004. Colour analysis of fruits using machine vision system for automatic sorting and grading. *Journal of Instrumentation Society of India* 34 (4), 284–291.
- Unay, D., Gosselin, B., Kleynen, O., Leemans, V., Destain, M., Debeir, O., 2011. Automatic grading of Bi-colored apples by multispectral machine vision. *Computers and Electronics in Agriculture* 75 (1), 204–212.
- Vidal, M., Amigo, J.M., Bro, R., Ostra, M., Ubide, C., Zuriarrain, J., 2011a. Flatbed scanners as a source of imaging. Brightness assessment and additives determination in a nickel electroplating bath. *Analytica Chimica Acta* 694 (1–2), 38–45.
- Vidal, M., Amigo, J.M., Bro, R., van den Berg, F., Ostra, M., Ubide, C., 2011b. Image analysis for maintenance of coating quality in nickel electroplating baths – real time control. *Analytica Chimica Acta* 706 (1), 1–7.

5.3. “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”, *Journal of Food Engineering* 153 (2015) 117-123 y “Methodology for marbling characterisation of pure Iberian cured loin based on hyperspectral imaging”. Comunicación oral 16th International Conference on Near Infrared Spectroscopy. 2013.





Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging



C. Garrido-Novell ^{a,*}, A. Garrido-Varo ^{a,1}, D. Pérez-Marín ^{a,1}, J.E. Guerrero-Ginel ^{a,1}, M. Kim ^{b,2}

^a University of Córdoba, Animal Production Department, Production System Engineering Unit, Córdoba 14014, Madrid national road, 396 km, Spain

^b United States Department of Agriculture, Sensing Technologies for Food Safety, Building 303 BARC-East, Room 001A Beltsville, MD 20705-2350, United States

ARTICLE INFO

Article history:

Received 11 July 2014

Received in revised form 13 September 2014

Accepted 22 September 2014

Available online 18 October 2014

Keywords:

NaCl
Moisture content
Hyperspectral
Validation
Textural analysis
Cured Iberian ham

ABSTRACT

Hyperspectral imaging technology is increasingly regarded as a powerful tool for the classification and spatial quantification of a wide range of agrofood product properties. Taking into account the difficulties involved in validating hyperspectral calibrations, the models constructed here proved moderately accurate for predicting NaCl content, while good accuracy was obtained for moisture content (Shenk and Westerhouse, 1996) in Iberian dry-cured ham slices. Using spatial information provided by chemical imaging, the distribution of both analytes was characterized by textural analysis as a function of muscle and commercial category.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Final salt levels in dry-cured ham vary considerably (from 4% to 10%, according to Arnau et al., 1994) both between muscles (depending on location, marbling, the salt and moisture permeability of the layer or layers covering the muscle, amongst other factors) and within the same muscle. This variability must be taken into account when attempting to reduce ham salt content, since irregular salt distribution detracts from the organoleptic and industrial quality of dry-cured ham (García-Gil et al., 2014).

Of the various technologies being developed at present for the indirect, fast, non-destructive measurement of NaCl in meat products, promising results have been reported for computerized tomography (Fulladosa et al., 2010) and Near Infra-Red Spectroscopy (NIRS) (Collell et al., 2011; Prevolnik et al., 2011).

NIR imaging technology records both spectral and spatial information in the form of an image, thus facilitating the quantitative and spatial characterization of heterogeneous samples (Gowen et al., 2008; Ravn et al., 2008). Techniques such as partial least

squares regression (PLS), principal component regression (PCR) and multilinear regression (MLR) can be applied to examine correlations between reference analyses and spectral data, enabling the creation of concentration maps for the component being analyzed (Burger and Geladi, 2006a).

In meat products, hyperspectral imaging has been used to detect contaminants, to characterize the technological and organoleptic quality of fresh and cured meat, and to classify commercial cuts. Kong et al. (2004) and Yang et al. (2006) used a combination of hyperspectral and fluorescence imaging to detect tumors and disease in poultry carcasses. Nakariyakul and Casasent (2008) and Park et al. (2007) detected fecal contamination in poultry carcasses using multispectral image analysis. Qiao et al. (2007a, 2007b) constructed models capable of characterizing fresh pork in terms of technological quality and marbling. Naganathan et al. (2008) used hyperspectral imaging to measure tenderness in beef, while Kamruzzaman et al. (2011) used this technique to distinguish between lamb cuts and muscles. ElMasry et al. (2011) applied hyperspectral imaging for the quality classification of cooked turkey hams. Barbin et al. (2011) graded and classified different fresh pork cuts as a function of technological quality.

Recent years have seen the development of applications for the pixel-based quantification of relevant chemical components in meat products. In 2012, Kamruzzaman et al. obtained PLS regression models for predicting moisture, fat and protein content in fresh lamb, with regression coefficients ranging from 0.63 for pro-

* Corresponding author. Tel.: +34 957212221.

E-mail addresses: g52ganoc@uco.es (C. Garrido-Novell), pa1gavaa@uco.es (A. Garrido-Varo), dcperrez@uco.es (D. Pérez-Marín), pa1gugij@uco.es (J.E. Guerrero-Ginel), moon.kim@ars.usda.gov (M. Kim).

¹ Tel.: +34 957212221.

² Tel.: +1 301 504 8450x245.

tein to 0.88 for moisture content, and a standard error of prediction (SEP) of 0.51% for moisture content. ElMasry et al. (2013) also used PLS models to measure moisture, fat and protein content in beef, obtaining a coefficient of determination of 0.89 and an SEP of 0.46% for moisture content. Iqbal et al. (2013) predicted moisture content, color and pH in cooked turkey ham slices, obtaining similar regression coefficients but slightly higher SEPs. In pork products, Barbin et al. (2013) and Talensa et al. (2013) predicted chemical composition (protein, moisture and fat) with a degree of success similar to that reported in other meat-product models.

The only papers focusing on the use of hyperspectral imaging to predict NaCl content in cured pork products would appear to be those published by Liu et al. (2013) and Gou et al. (2013). Liu et al. constructed MLR-based models to predict moisture and salt content in different cuts of fresh pork at different stages in the salting process, reporting coefficients of determination of 0.9 and 0.9 respectively, and SEPs of 0.682 for moisture and 0.007 for salt, using an instrument working in the range 400–1000 nm. Gou et al. (2013), using a more limited wavelength range to predict moisture and salt content, obtained coefficients of determination of 0.96 and 0.91 and SEPs of 1.34% and 0.71%, respectively.

One major problem reported in all attempts at pixel-based prediction using hyperspectral models is posed by the application of models constructed using mean reference and spectral value to the individual spectra from which they derive, with the attendant risk of artificiality or overfitting. The specific validation developed by Burger and Geladi (2006b) for hyperspectral prediction models tackles this issue.

This study sought to construct models for predicting NaCl and moisture content in Iberian dry-cured ham pixel by pixel, using NIRS imaging equipment (wavelength range 400–1700 nm) and the specific validation method developed by Burger and Geladi (2006b). Pixel-based predictions were subjected to textural analysis with a view to identifying patterns in NaCl and moisture content as a function of “commercial category” and “muscle”.

2. Materials and methods

2.1. Cured Iberian ham muscles

Sixty samples from sixty different dry-cured Iberian hams of varying quality classifications (20 acorn-fed pure Iberian -B-, 20 acorn-fed Iberian-Duroc crossbreds -Bx- and 20 feed-fed Iberian-Duroc crossbreds -C-) were studied. Each sample comprised 320 g of sliced ham taken transversely from the intermediate zone of each ham. Samples were purchased from two Iberian dry-curing industries in Córdoba, Spain. Before image acquisition, samples were removed from the fridge and kept for 30 min at room temperature (22 °C). Each intact sample was then imaged individually in the hyperspectral imaging system. Once hyperimages had been taken, each sample was divided into three subsamples as a function of muscular distribution (zone “SM”: *semimembranosus*, *abductor* and *gracilis* muscles, zone “BF”: *biceps femoris* muscle and zone “ST”: *semitendinosus* muscle). Subsamples were minced and homogenized. Moisture content was measured by oven drying to constant weight at 100 °C in accordance with ISO-R-1442, while NaCl concentration was determined by the Mohr method (UNE-ISO 9297:2013). A third of the samples were imaged in duplicate in order to calculate the Standard Error of Laboratory (SEL).

2.2. Image acquisition and processing

A line-scan hyperspectral imaging system operating in the NIR range was used. It comprised a XEVA-FPA-1.7–320-Xenic Charged Coupled Deviced (CCD) camera (resolution 320 × 256), a Specim

spectrophotometer (ImSpectr V10E), two 250 w halogen lamps and a stepper motor controlled platform (VXM, Velmex Inc., New York). During image acquisition, each sample was placed on the translation stage synchronously with the image acquisition in order to obtain a spectral image or hypercube with spatial resolution of $0.7 \times 0.8 \text{ mm}^2/\text{pixel}$. Each image was acquired in the spectral range 900–1700 nm with 3.15 nm intervals between contiguous bands and a total of 208 bands.

2.3. Multivariate analysis

Hyperspectral image regression involves the building of regression models between hyperspectral images ($n = 176$) and external reference data. Since external reference values may not be available for individual pixel locations, a single global reference value may be used for all the pixels in the hypercube. This single global value (reference data) must be related to a representative spectral signal of the hypercube, in this case the mean spectrum calculated from all the pixels of both faces of each muscle hyperimage in order to relate it to salt and moisture content (Burger and Geladi, 2006b).

Before building prediction models, outlier spectra were removed using Principal Component Analysis (PCA), Hotelling's T^2 (T-squared) and Q residuals with a 95% confidence level. After spectral outlier detection, the final sample set comprised 168 of the original 176 samples. Calibration equations were developed for each muscle from the mean spectra obtained by NIR-CI analysis using SIMPLS (De Jong, 1993). Classical PLS statistics were used for a preliminary evaluation of calibration equations.

Venetian-blind cross validation was applied to all data, drawing no distinction between calibration and prediction samples, as recommended by Burger and Geladi (2006b), in order to minimize the exaggerated imbalance between calibration samples and prediction samples (tens vs millions of millions). Models were validated using the following global statistics, specific for hyperspectral calibrations: global RMSEP (RMSEPG) and global Standard Deviation (SG) (Burger and Geladi, 2006b), where

$$\text{RMSEPG} = \left[\frac{\mathbf{r}^t \mathbf{r}}{P} \right]^{1/2} \quad (1)$$

$$\text{SG} = \left[\frac{\sum (s^2(L-1))}{\sum (L-1)} \right]^{1/2} \quad (2)$$

r: vector of bias values for each muscle.

P: the total number of hypercube muscles (176).

s: vector of standard deviation values for each muscle.

L: the number of pixels within each muscle in one image (not constant due to heterogeneous muscle shape and size).

SG represents the level of variance expected within one single image, while RMSEPG represents the accuracy expected for a new muscle prediction. In order to choose the best model, plots of RMSEPG versus SG were analyzed. For an underfit model RMSEPG may be quite large. As additional latent variables are added, the overall fit improves and RMSEPG is reduced. However, any additional variance included in the model that does not correlate with Y, will contribute to an increase in the variance of the predicted values, SG (Burger and Geladi, 2006b).

Two variability sources were included when constructing prediction models, with a view to their subsequent evaluation using the above statistics: spectral pretreatment and number of latent variables. Three different pretreatments and three different numbers of latent variables were used to build nine different models per constituent. The best model was selected in each case.

The Standard Normal Variate (SNV) transformation was chosen for scatter correction, and a Savitzky–Golay filter (derivative order:

1, polynomial order: 2, filter width: 15) (DER) for smoothing and as derivative treatment. The third option was the combination of the two, i.e. SNV + DER. Mean centering was the fourth type of pretreatment, applied in all cases to spectral and chemical data.

Although there is an optimum number of latent variables coinciding with the number of latent values, that minimizes the RMSECV without overfitting ($\text{RMSECV} < 1.3 \cdot \text{RMSEC}$, Shenk et al., 1989), two further values were chosen—one above and one below that optimum—in order to chart the response of RMSEPG and SG statistics to a change in the number of latent variables. The selection of these two additional latent variables was aimed at matching local minimum values for RMSECV without overfitting, as recommended by Shenk et al. (1989).

2.4. Histogram and textural analysis

In addition to individual chemical images, global histograms were created using NaCl and moisture content predictions from each pixel to detect general NaCl and moisture content distribution patterns. In order to obtain a parameter that summarized all the textural information contained in these images, a number of textural parameters were tested (Homogeneity, Second Angular Moment, Correlation and Contrast). Textural analysis was based on co-occurrence matrices, which examine the frequency with which pairs of neighboring values occur (Haralick et al., 1973).

Contrast (3) is an indicator of the contrast between neighboring pixel intensities across the whole image. A value of 0 is assigned to a constant image, and the value increases as image heterogeneity increases, in terms of pairs of values. Higher values for Homogeneity (5) are indicative of greater similarity of pixel pairs. The Angular Second Moment (6) is a measure of local homogeneity, i.e. it increases as the number of adjacent pixel pairs with exactly the same value increases. Correlation (4) is calculated differently, giving greater weight to pairs of values that deviate more from the average.

$$\sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \quad (3)$$

$$\sum_{i,j=0}^{N-1} P_{ij} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (4)$$

$$\sum_{i,j=0}^{N-1} P_{ij} / 1 + (i-j)^2 \quad (5)$$

$$\sum_{i,j=0}^{N-1} P_{ij}^2 \quad (6)$$

where i and j are salt or moisture concentrations and P_{ij} is the frequency of co-occurrence of i and j concentrations.

The method is summarized in Fig. 1.

3. Results and discussion

3.1. Construction of the hyperspectral prediction model

A critical step in the creation of accurate regression models is the choice of the optimal mathematical pretreatment and of the number of latent variables to use. This study used a validation method specifically designed for hyperspectral prediction models by Burger and Geladi (2006b). This method requires the prior construction of several models, from which one is chosen on the basis

of the global statistics SG and RMSEPG. For this purpose, 9 models were constructed, combining three different pretreatments and three different numbers of latent variables. All models complied with the requirement noted by Shenk et al. (1989) in order to avoid overfitting, i.e. $\text{RMSECV} < 1.3 \cdot \text{RMSEC}$.

The SEL for moisture content was 0.59%, while for NaCl content the SEL was 0.13 g/100 g. RMSECV values for NaCl content ranged between 0.64 and 0.97, in other words between 4.92 and 3.23 times the SEL; this, according to Shenk and Westerhaus (1996), is indicative of moderate accuracy. For moisture content, RMSECV ranged from 1.55 to 2.31, i.e. from 2.62 to 3.92 times the SEL, thus indicating a good degree of accuracy (Shenk and Westerhaus, 1996).

For all models except the NaCl prediction model constructed using the SNV + MC pretreatment, RMSEPG declined, and SG increased, as the number of latent variables increased (Figs. 2 and 3). When new latent variables were added, the overall adjustment (pooled predictions for all pixels) improved, giving rise to improved RMSEPG values. At the same time, the variance of predicted values increased, prompting a rise in SG. This increased variance is not indicative of greater real sample heterogeneity; rather, it is an artificial increase indicating that the model is less robust (Burger and Geladi, 2006b). The prediction error for each pixel increases, despite a decrease in the overall prediction error. The model chosen as optimal was therefore regarded as being the model which, whilst minimizing RMSEPG, did not give rise to an excessive SG. This can be done by defining a new parameter that combines both, or graphically by plotting RMSEPG vs SG. In a model with an optimal number of principal components, the distance from the plot origin to the points plotted for each parameter should be as small as possible (Burger and Geladi, 2006b).

Plots of RMSEPG vs SG for the nine models constructed to predict NaCl and moisture content are shown in Figs. 2 and 3, respectively; values for the model selected are encircled by a broken line. As Fig. 2 shows, though other models yielded lower values for RMSEPG, the DER + MC model (No. LV = 6) since it yielded a lower SG. Of the three pretreatments tested, the best results were obtained with the DER + MC combination.

The models DER + MC (No. LV = 7), DER + MC (No. LV = 12) and SNV + MC (No. LV = 12) yielded the lowest RMSEPG values for moisture content (Fig. 3). Of these, the lowest SG was obtained with DER + MC (No. LV = 7), encircled with a broken line.

The best model for predicting NaCl, as evaluated using the conventional RPD statistic, was the model constructed using 15 latent variables and the combined SNV y MC pretreatment (the same pretreatment yielded an RPD of 2.5 in Gou et al., 2013). However, as Fig. 2 shows, this model is the second furthest from the plot origin. The best RPD for moisture content was 3.23, lower than the 3.7 reported by Gou et al. (2013), also obtained using the SNV pretreatment. Nonetheless, the best overall model (Fig. 3) was DER + MC, 7.

3.2. Characterization of spatial distribution of salt and moisture content by muscle

Fig. 4 shows prediction maps for whole-slice samples (i.e. including all three muscles), grouped by quality category (from left to right, acorn-fed pure Iberian, acorn-fed Iberian-Duroc crossbreds and feed-fed Iberian-Duroc crossbreds). The upper row shows prediction maps for moisture content, and the lower row maps for NaCl content. In all three quality categories, the muscles with greatest moisture content are BF and ST, to the left of the maps. In the color scale used, warmer colors indicate greater moisture content. Dehydrated areas account for most of the SM muscle area (*semimembranosus*, *abductor* and *gracilis*) due to the absence of the protective subcutaneous fat layer found in the other two. Moisture content is less uniform in category C, where there is a clear

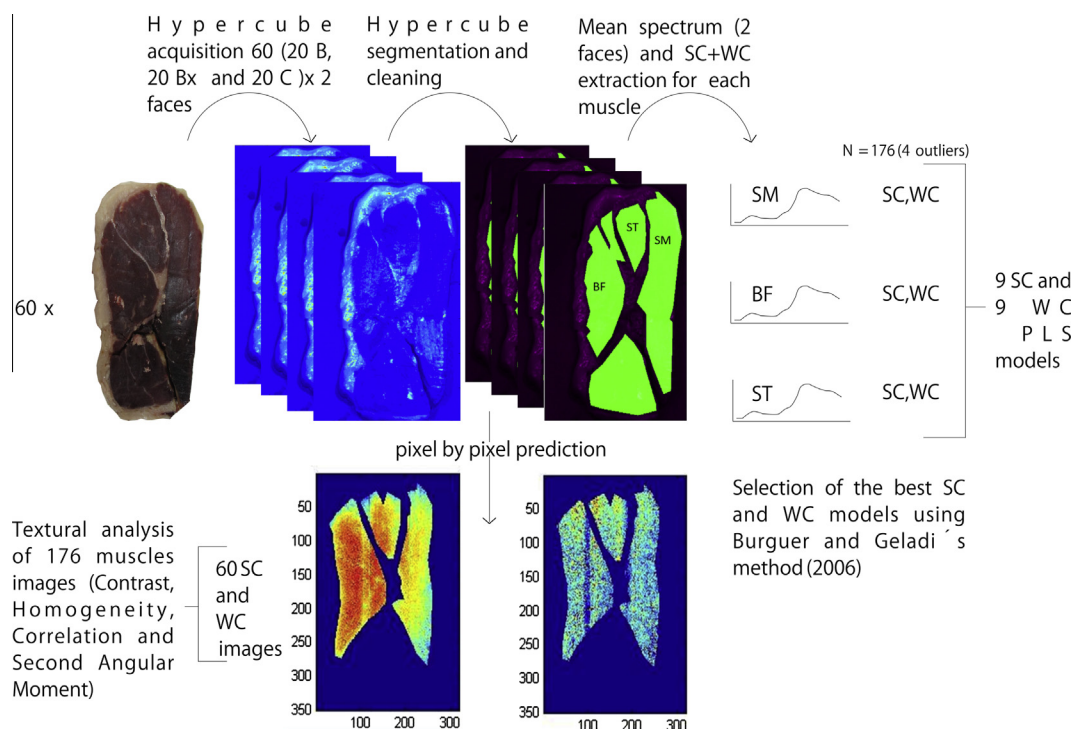


Fig. 1. General outline of the method (SC = salt content, WC = moisture content).

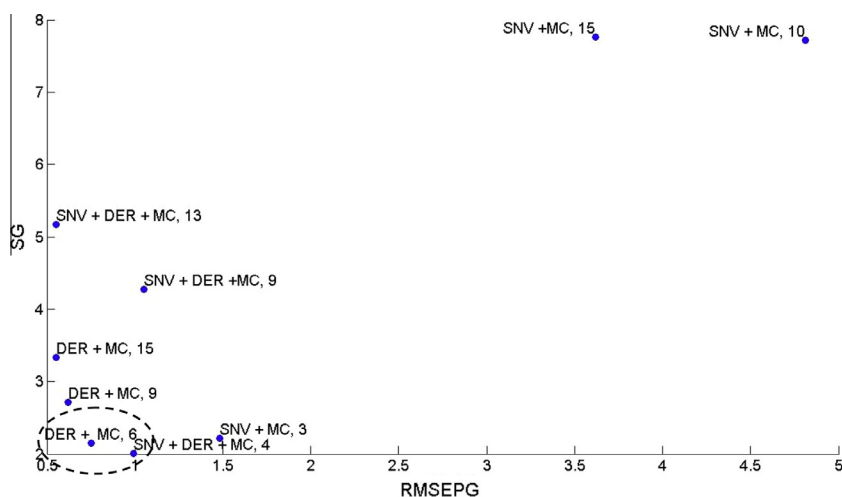


Fig. 2. RMSEPG vs SG for the nine models tested for predicting NaCl (g/100 g) in 176 prediction samples. The number refers to the number of latent variables used.

gradient from the most moist area (muscles *BF* and *ST*) to the most dehydrated area (*SM*). This effect of intra- and intermuscular moisture dynamics is less apparent in category Bx, and even less so in B.

Prediction maps for NaCl display far less intra-muscle heterogeneity, while inter-muscle differences, though apparent, are less marked than for moisture content.

Global frequency histograms constructed in order to characterize NaCl and moisture distribution by "Muscle" and "Category" at pixel level (all predicted pixels – $N > 4,500,000$ – for 176 samples) are shown in Fig. 5. Pixel frequency histograms were constructed using intervals of 0.7% for moisture content and 0.2 g/100 g for salt content. Pixel-based prediction revealed differences not only for the mean but also in frequency distribution.

As Fig. 5 shows, the frequency distribution curve for moisture content was much less pointed in category "C" than in the other

two categories, while muscles *BF* and *ST* displayed more marked left-side tails than muscle *SM*. Kurtosis and Fisher coefficients were calculated for all histograms (Table 1). Taking normal distribution as reference, only the distributions for categories "Bx" and "C" were platykurtic (i.e. less pointed and with narrower tails than the normal curve), while the rest were leptokurtic. The less pointed curve for moisture content in category C confirms the polarity appreciable in moisture content images for this category; indeed, the histogram is bimodal. This was also visible, albeit to a lesser extent, in BX samples, though not in B samples. The marked left-side tails of the distribution curves for moisture content in *BF* and *ST*, though displaying lower Fisher coefficient values than *SM*, were never negatively skewed. Skewness was positive in all histograms except for moisture content in "C", where the tail was more marked towards the right than towards the left. The

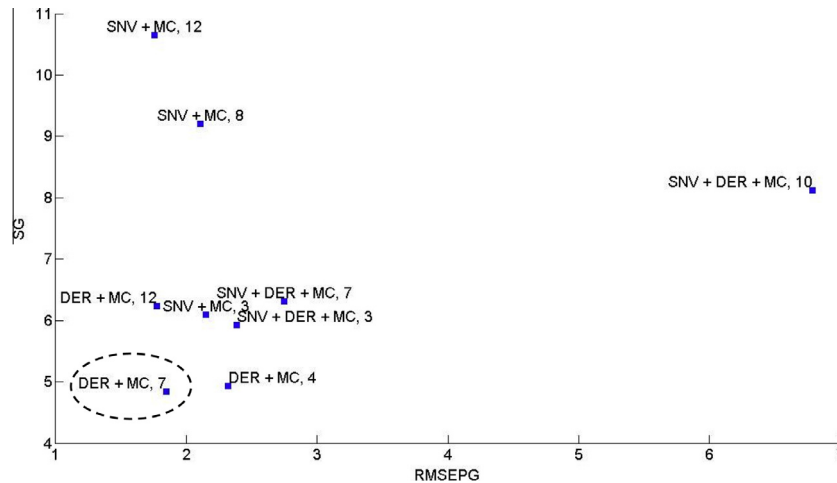


Fig. 3. RMSEPG vs SG for the nine models tested for predicting moisture content (%) in 176 prediction samples. The number refers to the number of latent variables and the letters to the mathematical pretreatment of the model (First Derivative – DER, Mean Center – MC, and Standard Normal Variate – SNV).

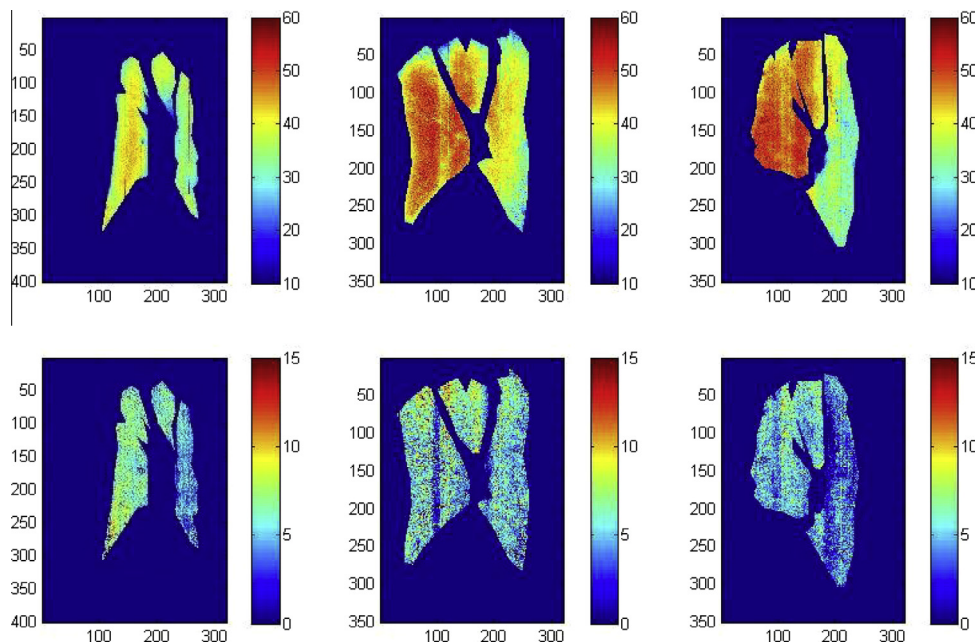


Fig. 4. Prediction maps for NaCl (lower row) and moisture content (upper row) for whole-slice samples (three muscle per sample) in each of the three categories. From left to right: acorn-fed pure Iberian, acorn-fed Iberian-Duroc crossbreds and feed-fed Iberian-Duroc crossbreds.

shape of the histograms for salt indicates only greater homogeneity between and within categories and muscles, thus generally confirming the findings on individual salt-content images.

In order to examine not only frequency but also spatial distribution, textural analysis was performed on the 176 images for NaCl and the 176 images for moisture content. Results (mean and standard deviation) for the textural features “Correlation”, “Contrast”, “Homogeneity” and “Second Angular Moment”, based on the calculation of co-occurrence matrices, are shown in Table 1. These features provide quantifiable spatial information to supplement the findings of visual analysis of individual images. The co-occurrence of pairs of values was examined horizontally (left and right), i.e. perpendicular to the surface through which NaCl enters and moisture is lost.

Significant inter-group differences were found for both the “Category” and the “Muscle” factors. Generally speaking, significant differences were more numerous for the “Muscle” factor

(affecting all textural features except the Second Angular Moment for moisture content). In the “Category” factor, certain combinations of textural features and analytes displayed no significant inter-group differences (e.g. “Correlation” for moisture content and both “Homogeneity” and “Second Angular Moment” for NaCl content).

“Correlation” increases as the repetition of pixel-pair patterns increases, i.e. when for the sample as a whole an increase or decrease in pixel i prompts a general increase or decrease in the neighboring pixel j . As a measure, it is thus equivalent to a gradient. For all groups and both factors, the horizontal gradient accounted for between 55% and 77% of local gradients or gradients between pixel pairs, the highest percentages being found for moisture content. The SM muscle displayed significantly lower values than ST and BF muscles both for NaCl and for moisture content. For the “Category” factor, only group B for NaCl content recorded significantly lower “Correlation” values than Bx and C. Both groups

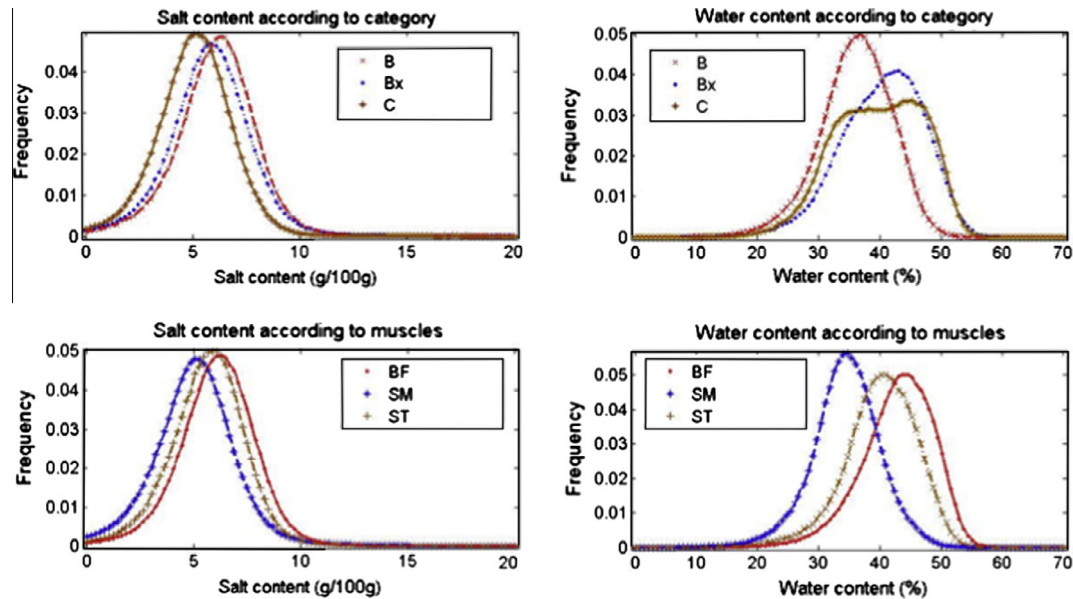


Fig. 5. Pixel frequency histograms for salt and moisture content, by muscle and category. BF: *biceps femoris*, SM: *semimembranosus*, ST: *semitendinosus*; B: acorn-fed pure Iberian, Bx: acorn-fed Iberian crossbred, C: feed-fed Iberian crossbred.

Table 1
Means and standard deviations for textural features “Correlation”, “Contrast”, “Homogeneity” and “Second Angular Moment” for NaCl and moisture prediction, by quality category and muscle.

	Correlation		Contrast		Homogeneity		Second Angular Moment	
	NaCl content	Moisture content	NaCl content	Moisture content	NaCl content	Moisture content	NaCl content	Moisture content
<i>Category</i>								
B	0.55 ^A ± 0.18	0.66 ^A ± 0.19	64.58 ^A ± 64.12	77.10 ^A ± 40.42	0.30 ^A ± 0.09	0.25 ^A ± 0.05	0.0048 ^A ± 0.0048	0.0018 ^{AB} ± 0.0012
Bx	0.62 ^B ± 0.15	0.68 ^A ± 0.16	42.96 ^B ± 34.77	61.76 ^A ± 29.94	0.31 ^A ± 0.07	0.27 ^B ± 0.05	0.0045 ^A ± 0.0032	0.0023 ^A ± 0.0020
C	0.66 ^B ± 0.12	0.71 ^A ± 0.12	40.81 ^B ± 32.68	77.24 ^A ± 33.90	0.31 ^A ± 0.06	0.24 ^A ± 0.03	0.0040 ^A ± 0.0026	0.0012 ^B ± 0.0004
<i>Muscles</i>								
BF	0.65 ^A ± 0.17	0.77 ^A ± 0.10	27.15 ^A ± 19.70	53.85 ^A ± 22.74	0.34 ^A ± 0.06	0.27 ^A ± 0.05	0.0058 ^A ± 0.0040	0.0018 ^A ± 0.0019
SM	0.56 ^B ± 0.13	0.58 ^B ± 0.17	42.65 ^A ± 23.40	79.33 ^B ± 35.94	0.32 ^A ± 0.07	0.26 ^A ± 0.05	0.0049 ^A ± 0.0039	0.0020 ^A ± 0.0015
ST	0.63 ^A ± 0.15	0.72 ^A ± 0.14	79.66 ^B ± 66.40	82.13 ^B ± 39.37	0.26 ^B ± 0.07	0.23 ^B ± 0.04	0.0028 ^B ± 0.0024	0.0014 ^A ± 0.0006

ANOVA ($\alpha = 0.05$). A and B superscripts indicate group membership according to Tukey test ($\alpha = 0.05$).

(B and SM) had the lowest moisture content for their respective factors (“Category” and “Muscle”, respectively). It may therefore be concluded that as moisture content increases, so does the percentage of the local gradient accounted for by the average gradient perpendicular to the analyte entry surface both for NaCl and for moisture content. Similar findings have been reported in studies involving a range of foods: the moisture diffusivity coefficient increases as moisture content increases (Okos et al., 1992).

The “Contrast” and “Homogeneity” features account for the same phenomenon in opposite ways. Whilst “Contrast” increases as the frequency of neighboring pairs displaying greater differences increases, “Homogeneity” increases as the frequency of differing pairs decreases. “Contrast” is therefore an indicator of local heterogeneity, while “Homogeneity” indicates local uniformity. The “Contrast” feature revealed significant inter-group differences for both analytes for both “Category” and “Muscle”, while the “Homogeneity” factor displayed no significant difference in means for the combination “Category” and NaCl, although of course opposing trends were apparent for the two features.

According to the “Contrast” feature, both NaCl and moisture content for the “Muscle” factor displayed the highest local heterogeneity for ST and the lowest for BF, while conversely the “Homogeneity” feature revealed the lowest local uniformity for ST. This may at first appear striking, since fat content and degree of curing are considered the major determinants of moisture and salt

content (Fulladosa et al., 2010; Håseth et al., 2012). In this respect, ST is more similar to BF than to SM. The explanation may be that ST is located between SM and BF. Consequently, as the images for NaCl and moisture content (Fig. 4) show, the gradients for moisture and, especially, for NaCl are much more marked, a finding confirmed by the higher values recorded for the “Contrast” feature for salt (Table 1).

The “Second Angular Moment” is greater when there are fewer large-magnitude entries, i.e. when there is a high degree of co-occurrence of the same pair of values; it is thus an indicator of singularities. The lowest value, i.e. the lowest number of singular values, was recorded for NaCl under the “Muscle” factor. Values for moisture content were significantly lower for group C than for Bx. Despite these differences, and at least under the assay conditions of the present study, this textural feature proved to be of little value for characterizing differences between muscles and quality categories, since it found no significant difference for the combinations “NaCl/Category” and “moisture content/Muscle”.

4. Conclusion

This study, using the specific validation method developed by Burger and Geladi (2006b) for hyperspectral prediction models, obtained prediction models of moderate accuracy for NaCl content and good accuracy for moisture content. It examined the spatial

and frequency distribution of salt and moisture content in Iberian dry-cured ham using a set of 176 samples and hyperspectral imaging equipment operating in the 900 a 1700 nm range.

Analysis of frequency histograms and textural analysis both proved to be useful tools for characterizing NaCl and moisture distribution on prediction maps. The textual feature “Correlation” indicated that the local gradient perpendicular to the surface of the ham in contact with the atmosphere increased as moisture content increased. The textural features “Homogeneity” and “Contrast” indicated that the local homogeneity of salt and moisture content was not governed by muscle fat content, and was not proportional to the depth of the muscle with respect to the surface of the ham in contact with the atmosphere; instead, it was greater in the *semitendinosus* muscle due to the intermediate position of this muscle between the surface in contact with the atmosphere and the surface covered by a layer of subcutaneous fat.

Acknowledgments

The authors acknowledge the financial support of the Excellence Project AGR 5129, “MEMS and NIRS-imaging sensors for non-destructive analysis of vegetable and animal products”. Cristóbal Garrido Novell acknowledges the financial support of the Spanish Ministry of Education as a fellow of the “University Teacher Training” Program (Formación del Profesorado Universitario, FPU) and the assistance of José Manuel Amigo and Juan Fernández Novales.

References

- Arnau, J., Guerrero, L., Casademont, G., Gou, P., 1994. Physical and chemical changes in different zones of normal and PSE dry cured ham during processing. *Food Chem.* 52, 62–69.
- Barbin, D., ElMasry, G., Sun, D.W., Allen, P., 2011. Near-infrared hyperspectral imaging for grading and classification of pork. *Meat Sci.* 90, 259–268.
- Barbin, D., ElMasry, G., Sun, D.W., Allen, P., 2013. Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chem.* 138 (2–3), 1162–1171.
- Burger, J., Geladi, P., 2006a. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst* 131, 1152–1160.
- Burger, J., Geladi, P., 2006b. Hyperspectral NIR image regression part II: dataset preprocessing diagnostics. *J. Chemom.* 20, 106–119.
- Collell, C., Gou, P., Arnau, J., Comaposada, J., 2011. Non-destructive estimation of moisture, moisture activity and NaCl at ham surface during resting and drying NIR spectroscopy. *Meat Sci.* 129, 601–607.
- De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* 18, 251.
- ElMasry, G., Iqbal, A., Sun, D.W., Allen, P., Ward, P., 2011. Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *J. Food Eng.* 103, 333–344.
- ElMasry, G., Sun, D.W., Allen, P., 2013. Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging. *J. Food Eng.* 117 (2), 235–246.
- Fulladosa, E., Santos-Garcés, E., Picouet, P., Gou, P., 2010. Prediction of salt and moisture content in dry-cured hams by computed tomography. *J. Food Eng.* 96, 80–85.
- García-Gil, N., Muñoz, I., Santos-Garcés, E., Arnau, J., Gou, P., 2014. Salt uptake and moisture loss in hams with different moisture contents at the lean surface and at different salting temperatures. *Meat Sci.* 96 (1), 65–72.
- Gou, P., Santos-Garcés, E., Høy, M., Wold, J.P., Liland, K.H., Fulladosa, E., 2013. Feasibility of NIR interactance hyperspectral imaging for on-line measurement of crude composition in vacuum packed dry-cured ham slices. *Meat Sci.* 95 (2), 250–255.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Bell, S.E.J., 2008. Recent applications of chemical imaging to pharmaceutical process monitoring and quality control. *Eur. J. Pharm. Biopharm.* 69, 10–22.
- Håseth, T.T., Sørheim, O., Høyb, M., Egelandsdal, B., 2012. Use of computed tomography to study raw ham properties and predict salt content and distribution during dry-cured ham production. *Meat Sci.* 90 (3), 858–864.
- Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Sys. Man Cybern.* SMC-3 (6), 610–621.
- Iqbal, A., Sun, D.W., Allen, P., 2013. Prediction of moisture, color and pH in cooked, pre-sliced turkey hams by NIR hyperspectral imaging system. *J. Food Eng.* 117 (1), 42–51.
- Kamruzzaman, M., ElMasry, G., Sun, D.W., Allen, P., 2011. Application of NIR hyperspectral imaging for discrimination of lamb muscles. *J. Food Eng.* 104, 332–340.
- Kamruzzaman, M., ElMasry, G., Sun, D.W., Allen, P., 2012. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. *Innovative Food Sci. Emerging Technol.* 16, 218–226.
- Kong, S.G., Chen, Y.R., Kim, I., Kim, M.S., 2004. Analysis of hyperspectral fluorescence images for poultry skin tumor inspection. *Appl. Opt.* 43 (4), 824–833.
- Liu, D., Qu, J., Sun, D.W., Pu, H., Zeng, X.A., 2013. Non-destructive prediction of salt contents and moisture activity of porcine meat slices by hyperspectral imaging in a salting process. *Innovative Food Sci. Emerging Technol.* 20, 316–323.
- Naganathan, G.K., Grimes, L.M., Subbiah, J., Calkins, J.R., Samal, A., Meyer, G.E., 2008. Visible/near-infrared hyperspectral imaging for beef tenderness prediction. *Comput. Electron. Agric.* 64, 225–233.
- Nakariyakul, S., Casasent, D., 2008. Hyperspectral waveband selection for contaminant detection on poultry carcasses. *Opt. Eng.* 47, 087202–087209.
- Okos, M.R., Narsimhan, G., Singh, R.K., Weitnauer, A.C., 1992. *Handbook of Food Engineering*, second ed. Marcel Dekker, New York, pp. 440.
- Park, B., Windham, W.R., Lawrence, K.C., Smith, D.P., 2007. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm. *Biosyst. Eng.* 96, 323–333.
- Prevolnik, M., Skrlap, M., Janes, L., Velikonja-Bolta, S., Skorjanc, D., Candek-Potokar, M., 2011. Accuracy of near infrared spectroscopy for prediction of chemical composition, salt content and free amino acids in dry-cured ham. *Meat Sci.* 88, 299–304.
- Qiao, J., Ngadi, M.O., Wang, N., Garipey, C., Prasher, S.O., 2007a. Pork quality and marbling level assessment using a hyperspectral imaging system. *J. Food Eng.* 83, 10–16.
- Qiao, J., Wang, N., Ngadi, M.O., Gunenc, A., Monroy, M., Garipey, C., 2007b. Prediction of drip-loss, pH and color for pork using a hyperspectral imaging technique. *Meat Sci.* 76, 1–8.
- Ravn, C., Skibsted, E., Bro, R., 2008. Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches. *J. Pharm. Biomed. Anal.* 48 (3), 554–561.
- Shenk, J.S., Westerhaus, M.O., 1996. Calibration de ISI way. In: Davies, A.M.C., Williams, P. (Eds.), *Near Infrared Spectroscopy: The Future Waves*. In *Proceedings of the Seventh International Conference on Near Infrared Spectroscopy*, UK.
- Shenk, J.S., Westerhaus, M.O., Abrams, S.M., 1989. Protocol for NIRS calibration: monitoring analysis results and recalibration. In: Marten, G.C., Shenk, J.S., Barten, F.E. (Eds.), *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*. USDA, Springfield, VA, pp. 104–110.
- Talensa, P., Mora, L., Morsy, N., Douglas, Barbin, F., ElMasry, G., Sun, D.W., 2013. Prediction of moisture and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging. *J. Food Eng.* (3), 272–280.
- Yang, C.C., Chao, K., Chen, Y.R., Kim, M.S., Early, H.L., 2006. Simple multispectral image analysis for systemically diseased chicken identification. *Trans. ASABE* 49 (1), 245–257.

Methodology for marbling characterisation of pure Iberian cured loin based on hyperspectral imaging.

Garrido-Novell C.^a; Pérez-Marín D C.^a; Garrido-Varo A.^a; Fernández-Novales J.^a; K. Moon S.^b; Guerrero J E.^a.

^aFaculty of Agriculture & Forestry Engineering University of Cordoba (Spain), Spain;

^bAgricultural Research Service, United States Departament of Agriculture (United States of America)

Corresponding author: crisgano8459@hotmail.com

Introduction

Marbling (amount and distribution of intramuscular fat streaks) of dry-cured Iberian pig (IP) products (hams, loins and sausages) is one of the most important factors in quality grading. It is usually performed by experts in a subjective way. Furthermore, fat streaks of dry-cured pure IP are less visible than those coming from the cross with the Duroc-Jersey breeds (Ventanas et al., 2007). The main goal of this paper is to develop and evaluate a methodology for an accurate and non-subjective characterisation of marbling properties of Iberian pig cured loins based on the information provided in a complete spectrum from each pixel taken from a spectral image.

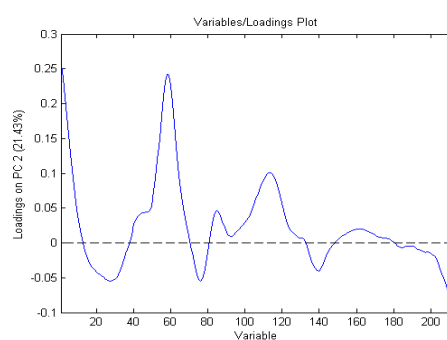
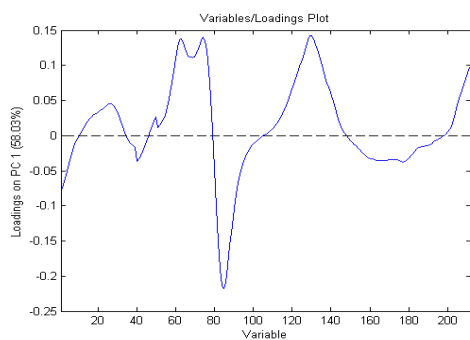
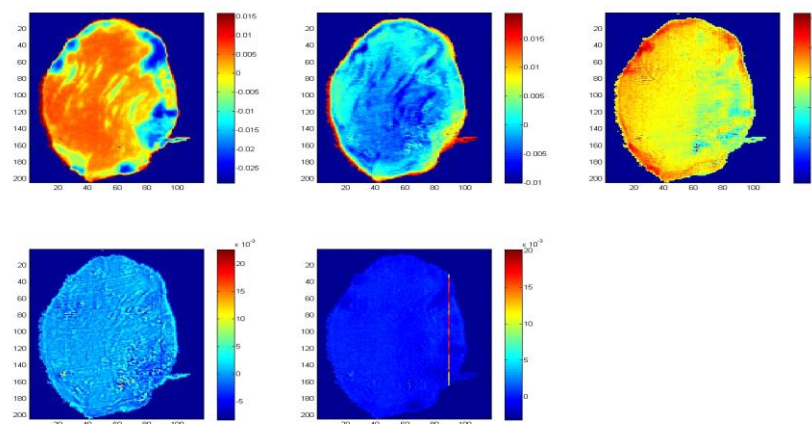
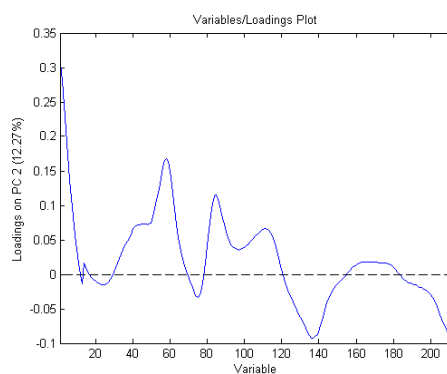
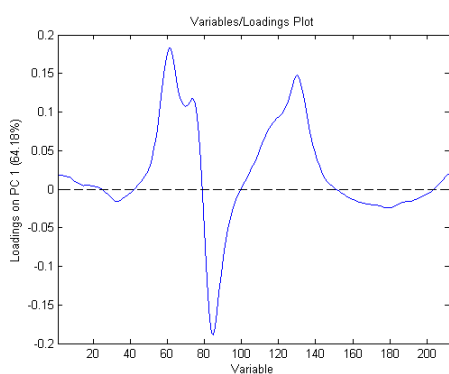
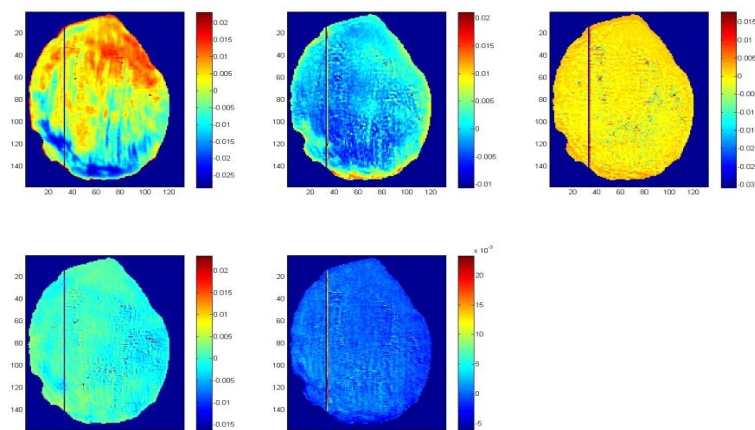
Material and methods

Nine transversal samples were obtained from the central part of 9 dry cured loins of different qualities. The first three samples came from three pure Iberian pigs fed with acorn (IB-A), the second three came from three cross-bred Iberian and Duroc- Jersey pigs fed with acorn (IBxD-A), and the last three came from Duroc Jersey breeds fed with concentrate feed (D-C). This selection of samples covered variability due to genetics (Pure Iberian, IberianxDuroc and Duroc) and feed (acorn or concentrate feed). Both RGB and spectral images were taken from each sample.

The spectral image acquisition system consisted of a Specim spectrometer (ImSpector V10E) for the NIR range (900-1700 nm), a pushbroom camera (XEVA –FPA-1.7–320-XeniCs CCD, resolution: 320x256) and two halogen lamps (500 w). Linear movement of the samples was performed by a step-motor platform (VXM, Velmex Inc., New York) coordinated with the camera. Both a white (99 % reflectance Spectralon; Labsphere, Inc., North Sutton, NH) and a black reference (lens closed) image were taken. Once reflectance had been calculated pixel-by-pixel (to avoid the introduction of noise due to different sensitivities between CCD pixels; Burger and Geladi, 2005), two background-sample segmentation methodologies were tried. In both cases, a threshold was applied over a monochrome image in order to mask all the channels with the resulting binary image. In the first

methodology, this monochrome image used the scores from a principal component analysis that best discriminated sample and background. In the other case it was a ratio between two channels. After that, different mathematical pretreatments were investigated such as multiplicative scatter correction (MSC), standard normal variate (SNV), autoscaling, mean centre (MC), first and second Savitzky-Golay derivative with different gaps and polynomial orders, smoothing of different polynomial orders and gaps, and their combination to minimise the influence of non-desired factors. Results were evaluated by—comparing the resultant scores from each pre-treated hyperimage; the choice of the best pre-treatment was made by a comparison of the variance explained by the first 5 principal components and the similarity between the first 5 scores and the RGB image (Fig. 1). For this purpose, principal component analysis was made over the sample spectra, once the background had been removed, considering each spectral image as a set of samples (Vidal and Amigo, 2012).

To determine fat level, a fuzzy cluster mean (FCM) model was used for each pretreated spectral image based on the complete acquired spectrum obtained on each pixel. FCM calculates the degree of membership of each spectrum to the mean spectrum of the two more important classes (Amigo, 2010). It is an unsupervised clustering model and so it was necessary to check that other components introduced during the dry cured loin processing, such as addition of dye, did not affect e.g. fat more than lean. Using the fat level of each spectrum from each pixel, fat level images (FLI) were created. In order to obtain a parameter that summarised all the textural information contained in these images, some parameters were investigated (Homogeneity, Second Angular Moment, Correlation and Contrast), and Contrast was finally chosen. A Contrast parameter was calculated from each FLI co-occurrence matrices (Qiao et al., 2007). Co-occurrence matrices show the frequency of each combination of pairs of contiguous values, in this case between 0 (lean) and 1 (fat). Contrast is an indicator of the contrast between neighbouring pixel intensity throughout the entire image. It is 0 for a constant image and increases with increasing image heterogeneity considering pairs of values (Fig. 2).



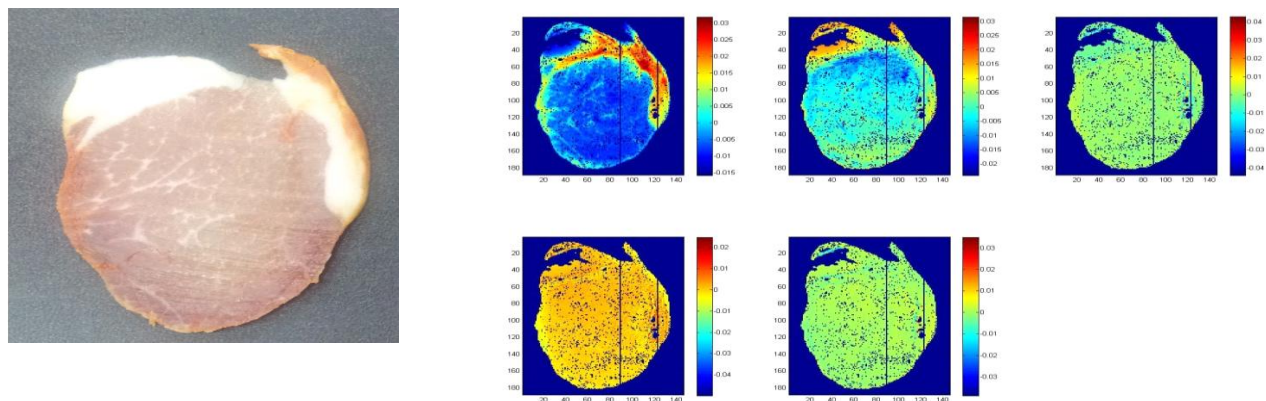


Figure 1. Five first principal component score images and loadings from RGB images of representative samples from the three categories analysed (PI-A, ID-A and D-C, from top to bottom).

$$\sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2$$

Figure 2. Local Contrast Formula.

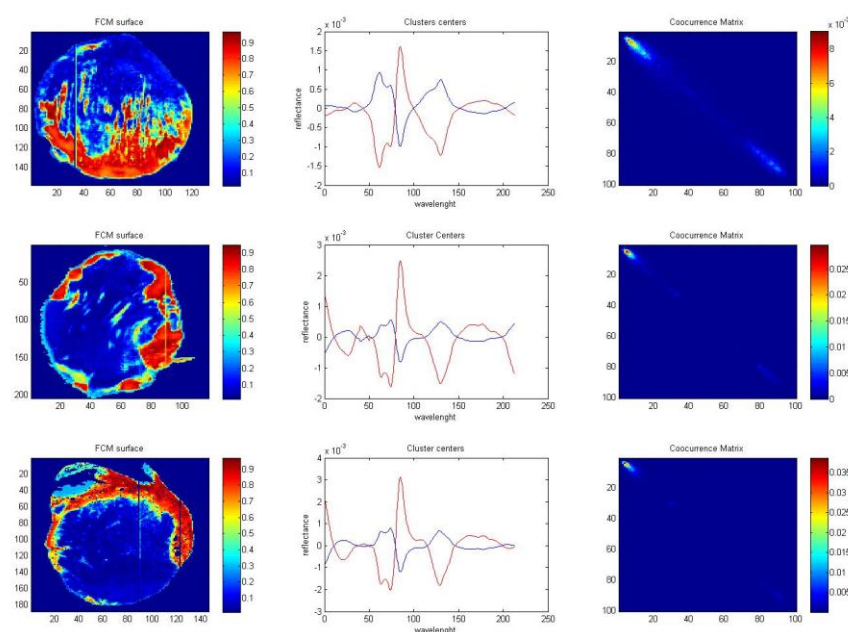


Figure 3. FLI, central spectra for each cluster and standardised co-occurrence matrix (from left to right) from the three analysed categories (PI-A, ID-A and D-C, from top to bottom).

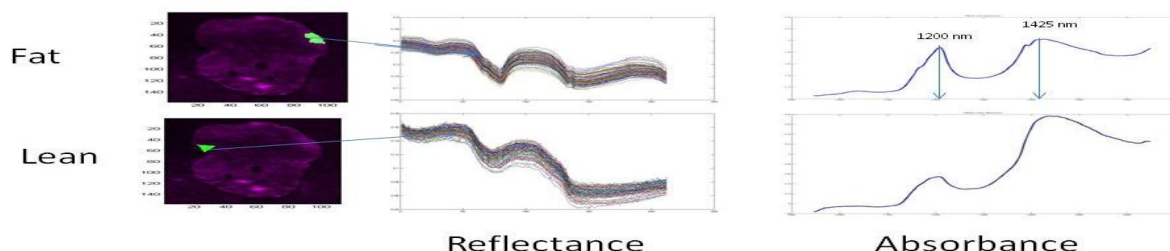


Figure 4. Extraction of known fat and lean spectra

Results and discussion

In Figure 1, the first five scores from three pre-treated and segmented spectral images (one of each category – PI-A, ID-A and D-C) can be seen. Score and ratio segmentation showed similar results (background consisted of a black plastic sheet). However, some pixels had to be removed manually in some images due to an effect of partial fat melting. The best pretreatment (Fig. 1) was obtained combining a MC and 1st derivative with a second order polynomial and 11 wavelength gap. Once noisy factors, such as scatter or brightness, have been removed by means of the pretreatment, 90 % of the explained variance is in the two first scores. Comparing the first score with the corresponding RGB image, it can be realised that they refer to the fat component (more or less fatty). It can be appreciated also in the loadings representation that the first or second component always refers to lean or fat.

However, scores values cannot be directly connected with a quantitative fat level. This ability to differentiate between levels of fat for each pixel based on the full spectrum is particularly important for the marbling characterisation in Iberian pig loins. Regardless of diet, Pure Iberian pigs have smaller, more irregular and weaker fat cells than the crossbred (IB x D), usually provoking rupture and spreading of melted fat on the meat surface. The goal of using the FCM model was to obtain a quantitative prediction of each pixel fat level (degree of membership to fat cluster). FLI from the same three representative samples can be seen in Figure 3 (red: fat, blue: lean).

Both the central spectrum of two main classes and the standardised co-occurrence matrix of each FLI may be discerned. It can be seen that the central spectrum of each class, seen in FLI and corresponding RGB images, corresponds to fat and lean pixels in the limits, in agreement with sample loadings in Figure 1.

However, in order to ensure that both clusters referred to fat and lean, known fat and lean spectra from every category were taken and pre-treated in exactly the same way as the modeled ones (Fig. 4). Lean and fat pre-treated spectra can be seen (top left in Figure 5). It can be appreciated how the shape of these spectra agree with both clusters (fat and lean) of the central spectra for IB, IBxD and D.

Co-occurrence matrices show the frequency of each contiguous pair of values (from 0 lean to 1 fat in 100 parts), producing a representation of fat level distribution. So, a concentration of high values (higher frequencies) in the diagonal indicates a high frequency of pairs of similar values, while a higher weight of one end of the diagonal or the other indicates a higher frequency of high (next to 1) or low (next to 0) fat value pairs; a higher weight of its centre indicates a higher co-occurrence of values close to 0.5. Co-occurrence matrices were calculated for all pairs of pixel orientation, that is 0°, 45°, 90° and 135°. The 0° co-occurrence matrices are shown in Figure 3.

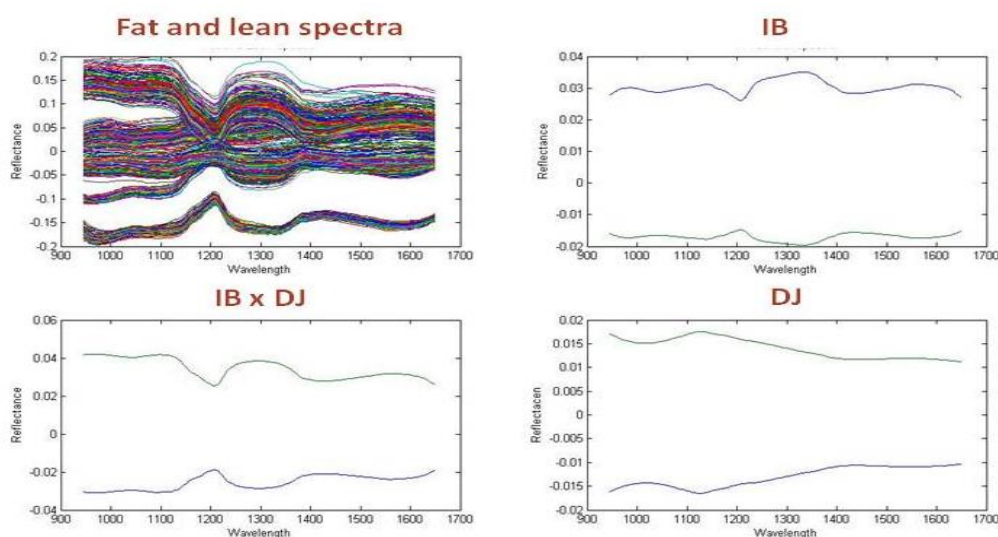


Figure 5. Spectral shape comparison between known pre-treated fat and lean spectra and central classes for each category given by FCM

It is necessary to summarise the information provided by co-occurrence matrices in a parameter that gathers the information of the entire matrix (distribution of fat levels) as could be done for a monochrome image. Figure 6 shows contrast parameter values for each sample for different pairs of pixel orientations. Although no clear trend can be seen, if the mean contrast values for each group (PI-A, ID-A and D-C) are drawn, a downward trend can be detected (Figure 7).

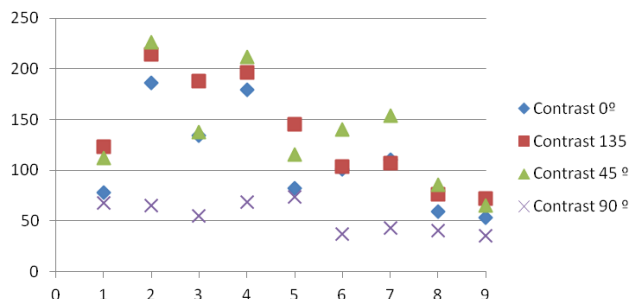


Figure 6. Contrast parameter versus samples (1-3, PI-A, 4-6, ID-A, and 7-9 D-C).

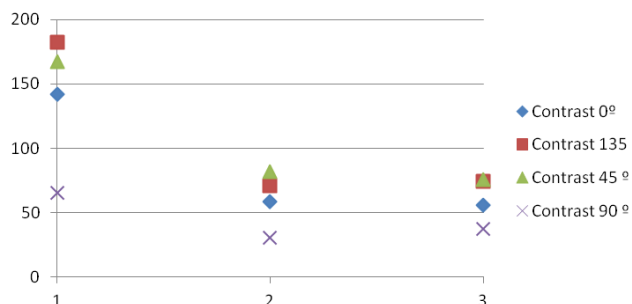


Figure 7. Mean contrast parameter versus groups (1, PI-A, 2, ID-A and 3 D-C).

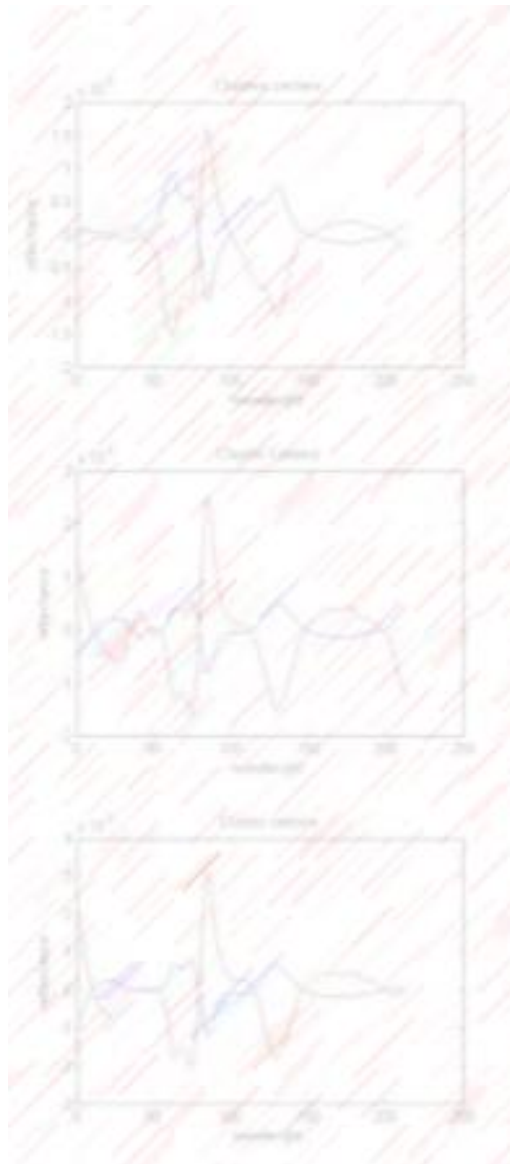
Conclusion

This study showed that FLI based on FCM and co-occurrence matrix analysis applied to hyperspectral imaging is a promising tool to predict fat level and fat distribution in Iberian pig cured loin. The preliminary results obtained are the basis for deeper research undertaken as part of the PhD research of the first author. Some points to improve are optimisation of cluster size, wavelength range and library size and representativeness.

References

- Amigo, J. M. (2010). Practical Issues of Hyperspectral Imaging Analysis of Solid Dosage Forms. *Anal. Bioanal. Chem.*, **398**(1), 93-109.
- Burger, J. and Geladi, P. (2005). Hyperspectral NIR image regression part I: Calibration and correction. *J. Chemometr.*, **19**, 355-363.
- Qiao, J., Ngadi, M. O., Wang, N., Gariépy, C., Prasher, Shiv. O. (2007). Pork quality and marbling level assessment using a hyperspectral imaging system. *J. Food Eng.*, **83**, 10-16.
- Ventanas, S., Ventanas, J. and Ruiz, J. (2007). Sensory characteristics of Iberian dry-cured loins: Influence of cross-breeding and rearing system. *Meat Sci.*, **75**, 211-219.
- Vidal, M. and Amigo, J.M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemomet. Intell. Lab. Sys.*, **117**, 138-148

5.4. "Differentiation of beef and fish meals in animal feeds using chemometric analytic models", *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)



Differentiation of Beef and Fish Meals in Animal Feeds Using Chemometric Analytic Models

Chun-Chieh Yang¹, Cristóbal Garrido-Novell², Dolores Pérez-Marín², José E Guerrero-Ginel²,
Ana Garrido-Varo², Hyunjeong Cho³, Moon S. Kim^{1*}

¹*Environmental Microbial and Food Safety Laboratory, Building 303 BARC-East, USDA-ARS,
10300 Baltimore Avenue, Beltsville MD 20705 USA*

²*Faculty of Agriculture and Forestry Engineering, Non-Destructive Spectral Sensors Unit, University of Cordoba, Spain*

³*Experiment & Research Institute, National Agricultural Products Quality Management Service, Republic of Korea*

Received: May 13rd, 2015; Revised: May 28th, 2015; Accepted: May 30th, 2015

Abstract

Purpose: The research presented in this paper applied the chemometric analysis to the near-infrared spectral data from line-scanned hyperspectral images of beef and fish meals in animal feeds. The chemometric statistical models were developed to distinguish beef meals from fish ones. **Methods:** The meal samples of 40 fish meals and 15 beef meals were line-scanned to obtain hyperspectral images. The spectral data were retrieved from each of 3600 pixels in the Region of Interest (ROI) of every sample image. The wavebands spanning 969 nm to 1551 nm (across 176 spectral bands) were selected for chemometric analysis. The partial least squares regression (PLSR) and the principal component analysis (PCA) methods of the chemometric analysis were applied to the model development. The purpose of the models was to correctly classify as many beef pixels as possible while misclassified fish pixels in an acceptable amount. **Results:** The results showed that the success classification rates were 97.9% for beef samples and 99.4% for fish samples by the PLSR model, and 85.1% for beef samples and 88.2% for fish samples by the PCA model. **Conclusion:** The chemometric analysis-based PLSR and PCA models for the hyperspectral image analysis could differentiate beef meals from fish ones in animal feeds.

Keywords: Beef, Chemometrics, Fish, Hyperspectral image, Line-scan, NIR

Introduction

European Union (EU) regulated zero tolerance for the existence of meat and bone matters (MBM), as part of animal protein sources, in compound animal feeds to respond to the increasing demands to protect and ensure food safety and public health (Riccioli et al., 2011). The existence of banned matters in animal feeds could cause serious diseases, such as the bovine spongiform encephalopathy (BSE). To satisfy the demands from the regulations and the public health concerns, it is essential to automatically identify and differentiate the contents of

animal feeds in the animal-feeds processing plants (Brookes, 2001; Riccioli et al., 2011). The non-destructive technology using machine vision for such automatic detection can be potentially the optimal promising solution to satisfy the demands of food safety and public health (He et al., 2014; June et al., 2013; Lee et al., 2014; Yang et al., 2014).

Among various non-destructive technologies of optical engineering, the hyperspectral imaging has been applied for food safety and quality purposes (Elmasry et al., 2012; He et al., 2014; June et al., 2013; Lee et al., 2014; Yang et al., 2010; 2014). Specifically, the hyperspectral imaging technology has been used in the attribute analysis for many animal food products, such as fish (He et al., 2014), beef (Elmasry et al., 2012), and poultry (Yang et al., 2010). To apply the hyperspectral imaging technology to the

*Corresponding author: Moon S. Kim

Tel: +1-301-504-8450; Fax: +1-301-504-9466

E-mail: moon.kim@ars.usda.gov

automatic detection in the food product processing line, line-scanning machine vision system has been developed and shown great potential for quick and reliable food safety inspection (Yang et al., 2010; 2014).

On the other hand, hyperspectral imaging in the line-scanning machine vision system could continuously and quickly produce huge quantity of spectral and spatial data for image analysis and classification. To quickly and properly analyze such a huge quantity of hyperspectral image data, chemometric analysis could be a reliable statistical tool to generating correlation models (Garrido-Novell et al., 2015). Therefore, the hyperspectral line-scanning machine vision system implemented with the chemometric-analysis-based model will be very useful and essential for the online application of food safety and quality inspection.

The fish meal is a high protein source of the compound animal feeds. It is important to detect banned animal matters, such as beef, in fish meals. Therefore, the primary object of this research was to apply the chemometric analysis to the development of the classification models using hyperspectral data to distinguish beef and fish meals in animal feeds. Each pixel of the Region of Interest (ROI) of the meal image was treated as an independent object for model development and evaluation. The main goal of the first step of the research reported in this paper was to correctly classify as many pixels of beef meals as possible for public health concern.

Materials and Methods

This research collected 15 samples of beef meals and 40 samples of fish meals prepared in a rendering plant. The meal samples could randomly contain various materials, such as muscle, skin, blood, bone, grease, and feather. One gram of each rendered meal sample was retrieved and carefully spread in a black paper for line-scanning. The black paper was used as a dark background. The hyperspectral line-scan machine vision system was composed of a XEVA-1.7-320 CCD camera (Xenics, Leuven, Belgium), an ImSpector V10E spectrograph (Specim, Oulu, Finland), two 250-W halogen lights and a VXM stepper motor controlled platform (Velmex, Bloomfield, NY) (Garrido-Novell et al., 2015). The original size of an image was 320×256 pixels. The image resolution was approximately $0.5 \text{ mm}^2/\text{pixel}$. The camera and the spectrograph were built specifically for the near infrared spectral range, in which

the hyperspectral data would present more useful information for proper classification of beef and fish meals.

While scanning the samples, they were placed on the platform and moved through the field of view of the camera to obtain line-scan images. Each image represented a scanned line along the field of view of the camera. Through the slit of the spectrograph, incoming light was dispensed into a spectrum for each pixel of the line image. Therefore, a line-scanned image contained spectral data along one and spatial data along another axis.

In this research, an original line-scanned image contained spectral data along one axis of 320 pixels, with a 3.3-nm spectral interval, and spatial data along another axis of 256 pixels. The trial-and-error-based preliminary study showed the need to discard 144 wavebands where the raw account of light reflectance from the scanned objects was too low or too noisy. Therefore, only 176 spectral-dimension pixels spanning from 969 nm to 1551 nm were used for data analysis and model development. Five meal samples were put on the platform and moved through the field of view of the camera to be scanned for 200 times.

The trial-and-error-based preliminary study also showed the need to identify a ROI in order to discard non-meal, dark background or improperly-imaged-meal. The size of ROI was manually selected to cover 3600 pixels for each sample. The images of eight beef meals and 20 fish meals, the development group, were used for hyperspectral data analysis and chemometric model development. The development group contained 28800 pixels for beef meals and 72000 pixels for fish meals. Each pixel would be an independent hyperspectral data source and differentiation object. The developed models were then applied to the images of the rest seven beef meals and 20 fish meals, the test group. The test group contained 25200 pixels for beef meals and 72000 pixels for fish meals. The models were evaluated using the results from the test group for the optimal model attributes.

After images were acquired, the line-scan images were converted from raw account to relative reflectance by the flat-field calibration with a reference white line image, W , and a reference dark current line image, D . Before meal samples were scanned, a white target with 99% diffuse reflectance, moving on the motor platform through the field of view of the camera, was scanned for four times. The acquired white line images were averaged to form the reference white line image. Afterward, the lens of the camera was covered in order to block all illumination

from the environment to take another four shots. The acquired dark line images were averaged to form the reference dark current line image. With the averaged reference images W and D , the raw hyperspectral images (I_0) of meal samples were calibrated and converted to the relative reflectance images (I) by the equation 1:

$$I = \frac{I_0 - D}{W - D} \quad (1)$$

The examples of calibrated images of beef and fish meal samples were shown in Figure 1. The intensity of the ROI in each image was increased intentionally for locating and viewing only. Within the ROI of five beef and five fish meal sample images randomly selected from the development group, the averaged spectra and the ranges of one standard deviation were calculated and shown in Figure 2. The hyperspectral graph in Figure 2 showed two highest spectral peak tops. The one in the shorter waveband was in the waveband of 1132 nm for beef and the top was slightly shifted to 1126 nm for fish. The one in the

longer waveband was in the waveband of 1302 nm. In Figure 2, the reflectance for beef meals was generally higher than one for fish in wavebands shorter than 1302 nm. Specifically, the spectral difference between beef and fish meals could be observed easily in the spectral peak from 969 nm to 1192 nm. The spectral difference between beef and fish in 1132 nm, one of the spectral peak top, was much higher than that in 1302 nm, another spectral peak top. Such difference could be observed in Figure 1. On the other hand, the spectral difference was relatively low for the spectral range from 1192 nm to 1551 nm.

Although the spectral difference in Figure 2 was significant, such difference could not be used for the proper classification of beef and fish meals because the spectral variance was so high to possibly cause low success differentiation rate. High spectral variance was expected because the spectral reflectance from the meal samples could come from muscle, skin, blood, bone, grease, feather, or other materials of the rendered animal. To overcome this challenge, the chemometric analysis, a computationally fast statistical method ideally for complicated data modeling,

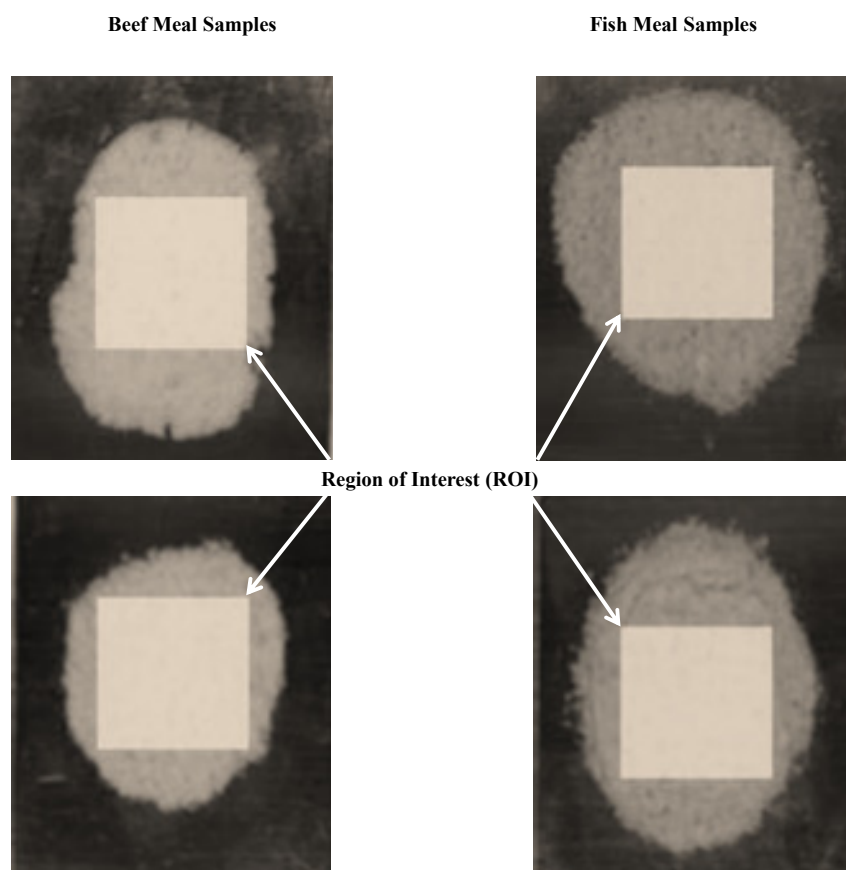


Figure 1. The samples for the line-scanned images of beef meals and fish meals with the Region of Interest (ROI) highlighted.

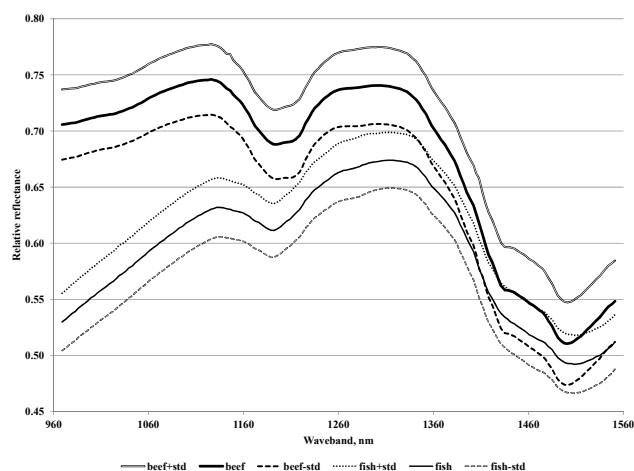


Figure 2. The spectra, in relative reflectance, for the hyperspectral images of beef meals and fish meals: the average spectrum plus one standard deviation (beef+std, fish+std), the average spectrum (beef, fish), and the average minus one standard deviation (beef-std, fish-std).

was applied to the development of the differentiation model.

Two most popularly used statistical tools of the chemometric analysis, the partial least squares regression (PLSR) and the principal component analysis (PCA), were applied to the model development. Both tools were reliable and robust for multivariate data analysis and modeling by generating components as new input variables to linearly compose original input variables (the relative reflectance in each waveband for this experiment). The components required for accurate model performance was usually much fewer than the original input variables. While generating the set of components to reduce dimension of the input variables, the PLSR intended to consider the impacts of outputs but the PCA would search the variance among the input variables. The details for the calculation and comparison of the PLSR and PCA can be referred to Godoy et al. (2014), Westerhuis et al. (1998) and Worley et al. (2013).

The models based on each tool were generated with components varied from one to ten for the optimal algorithm. The model required the relative reflectance from 176 wavebands from a pixel of the hyperspectral image for the inputs. The binary model output would determine the matter in this pixel to be beef or fish meal: zero for fish and one for beef. The total 100800 pixels of the development group were used for model development, and the total 97200 pixels of the test group were used for model evaluation. The models were developed, executed, and evaluated on the platform of MATLAB R2013a

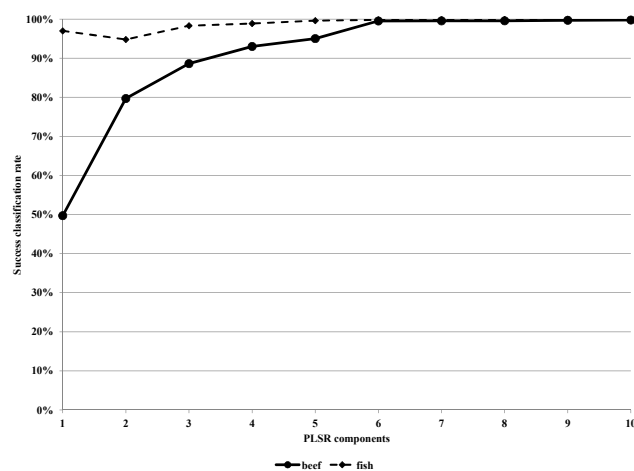


Figure 3. The results of the success classification rates for the PLSR models on the development group of 28800 pixels for eight beef meals and 72000 pixels for 20 fish meals.

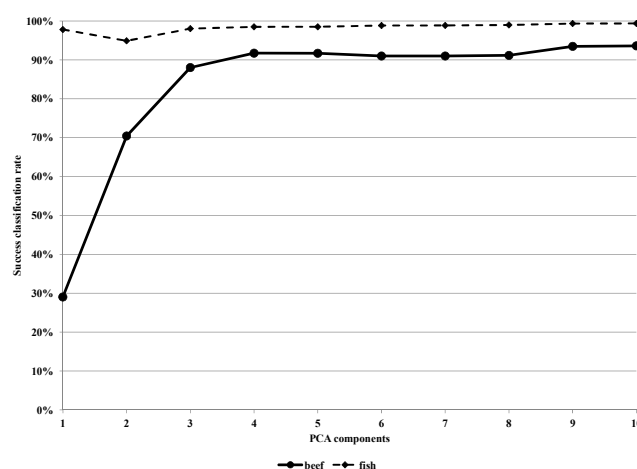


Figure 4. The results of the success classification rates for the PCA models on the development group of 28800 pixels for eight beef meals and 72000 pixels for 20 fish meals.

(MathWorks Inc., Natick, MA, USA).

Results and Discussion

The results in Figure 3 showed that the PLSR models were generally developed well for the classification of beef and fish meals. With at least six components, the PLSR models could correctly classify more than 99% of pixels for both beef meals (28800 pixels in total) and fish meals (72000 pixels in total). The PLSR models misclassified less than 140 pixels for both beef and fish meals with at least six components, and even missed less than 100 pixels with at least nine components. The model performance was generally improved with more components during the model development.

The results in Figure 4 showed the similar well development for the PCA models. With at least four components, the PCA models could correctly classify more than 90% of pixels for beef meals and more than 98% of pixels for fish meals. Comparing Figure 4 with Figure 3, the PCA models could not perform as well as the PLSR models did during the model development. Regardless to the number of the components, the PCA models misclassified more than 1800 pixels for beef meals and more than 450 pixels for fish meals. However, less components could require less computational effort for the model to classify the pixels, which would be essential for online application for high-speed processing line. Because of increasing concerns for food safety, the possible economic loss caused by misclassification of fish meals in Figures 3 and 4 might be tolerable. It is more important to accurately detect beef matters due to zero tolerance for beef matters in fish meals; thus, the PLSR models performed much better than the PCA models during the model development. Nevertheless, both PLSR and PCA models showed significant potential for further analysis to differentiate beef matters from fish meals.

Figure 5 showed the PLSR models performed on the test group of beef meals (25200 pixels in total) and fish meals (72000 pixels in total). Similar to Figure 3, the PLSR models performed well with at least six components when the success classification rate was also higher than 99% of pixels for fish meals, and could correctly classify more than 96% of pixels for beef meals. The highest success classification rates were obtained by the PLSR model with nine components, 97.9% of pixels for beef meals

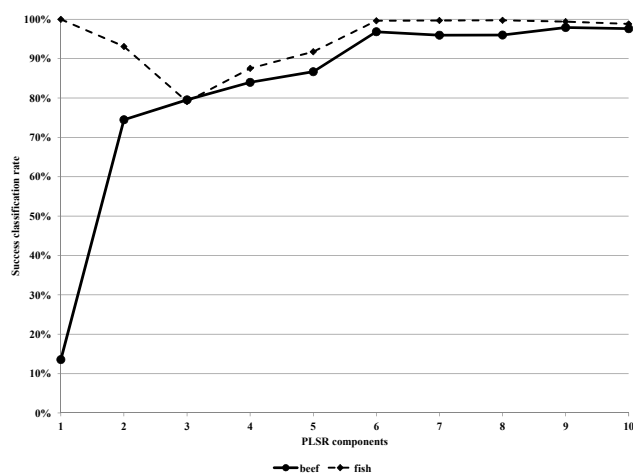


Figure 5. The results of the success classification rates for the PLSR models on the test group of 25200 pixels for seven beef meals and 72000 pixels for 20 fish meals.

(527 pixels misclassified) and 99.4% of pixels for fish meals (420 pixels misclassified). The high success classification rate for the pixels independent from the ones used for model development indicated that the PLSR models could be used in the real online application for food safety. The PLSR model would be built with six components when the computation speed is more essential, or with nine components when the most accurate identification of beef matters is required to exchange the computation speed.

Different from Figure 5, the results in Figure 6 showed that the PCA models could not successfully classify more than 90% of pixels for both beef and fish meals with the same model algorithm. With at least four components, the PCA models would successfully classify more than 81% of pixels for beef meals and more than 86% of pixels for fish meals. Although the results in Figure 6 was not as satisfactory as ones in Figure 5, the high success classification rates in Figure 4 implied that the performance of the PCA models could be improved with more data presented to the development group. This possible improvement could also be applied to the PLSR models.

Moreover, the samples were not scanned on the same day. Fifteen samples of beef meals were scanned on three days and 40 samples of fish meals were scanned on eight days, respectively. Considering the possible variance in the imaging environment and labor effort, the high success classification rates in Figures 3 to 6 showed that both PLSR and PCA models could perform well for time-series data. It is essential since the models were expected to be applied to the real online application for the future

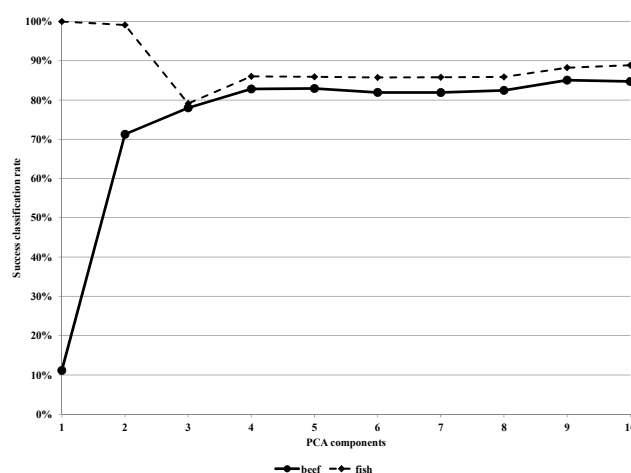


Figure 6. The results of the success classification rates for the PCA models on the test group of 25200 pixels for seven beef meals and 72000 pixels for 20 fish meals.

development.

The results showed the great potential applying the chemometric analysis-based algorithms, such as the PLSR and the PCA models, to the hyperspectral image analysis in order to differentiate beef meals from fish ones in animal feeds. The well-developed chemometric models could potentially handle high spectral variance caused by different materials of the rendered animal in the same meal. To further improve the model performance, more images would be collected for model development. Also, more proper illumination scheme would be designed for the line-scan machine vision system.

Conclusion

This research reported the development and the evaluation of the chemometric analysis-based models using the partial least squares regression (PLSR) and the principal component analysis (PCA) methods for the differentiation of beef and fish meals in the line-scanned hyperspectral images. The hyperspectral data from 176 wavebands spanning from 969 nm to 1551 nm were used for model inputs. With nine components in the model algorithm, the success classification rates were 97.9% for beef meals and 99.4% for fish meals by the PLSR-based model, and 85.1% for beef meals and 88.2% for fish meals by the PCA-based model. High success classification rates indicated the greatly potential application of the hyperspectral line-scan machine vision system implemented with the chemometric analysis model for the detection of beef matters in fish meals. The system would provide essential help to ensure food safety and public health.

Conflict of Interest

The authors have no conflicting financial or other interests.

References

Brookes, G. 2001. The EU animal feed sector: protein ingredient use and implications of the ban on use of meat and bonemeal. Brookes West, Jasmine House, Kent, UK.
Elmasry, G., D. F. Barbin, D. W. Sun and P. Allen. 2012.

Meat quality evaluation by hyperspectral imaging technique: an overview. *Critical Reviews in Food Science and Nutrition* 52(8):689-711.

Garrido-Novell, C., A. Garrido-Varo, D. Pérez-Marín, J. E. Guerrero-Ginel and M. S. Kim. 2015. Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging. *Journal of Food Engineering* 153:117-123.

Godoy, J. L., J. R. Vega and J. L. Marchetti. 2014. Relationships between PCA and PLS-regression. *Chemometrics and Intelligent Laboratory Systems* 130:182-191.

He, H.-J., D. Wu and D.-W. Sun. 2014. Potential of hyperspectral imaging combined with chemometric analysis for assessing and visualising tenderness distribution in raw farmed salmon fillets. *Journal of Food Engineering* 126:156-164.

Jung, D., M. S. Kim, K. Chao, H. Lee, H. Y. Lee and B. Cho. 2013. Detection algorithm for cracks on the surface of tomatoes using multispectral Vis/NIR reflectance imagery. *Journal of Biosystems Engineering* 38(3):199-207.

Lee, H. Y., C.-C. Yang, M. S. Kim, J. Lim, B. Cho, K. Chao, A. Lefcourt and C. Everard. 2014. A simple multispectral imaging algorithm for detection of defects on Red Delicious apples. *Journal of Biosystems Engineering* 39(2):142-149.

Riccioli, C., D. Pérez-Marín, J. E. Guerrero-Ginel, J. E. Guerrero-Gienl and A. Garrido-Varo. 2011. Pixel selection for near-infrared chemical imaging (NIR-CI) discrimination between fish and terrestrial animal species in animal protein by-product meals. *Applied Spectroscopy* 65(7): 771-781.

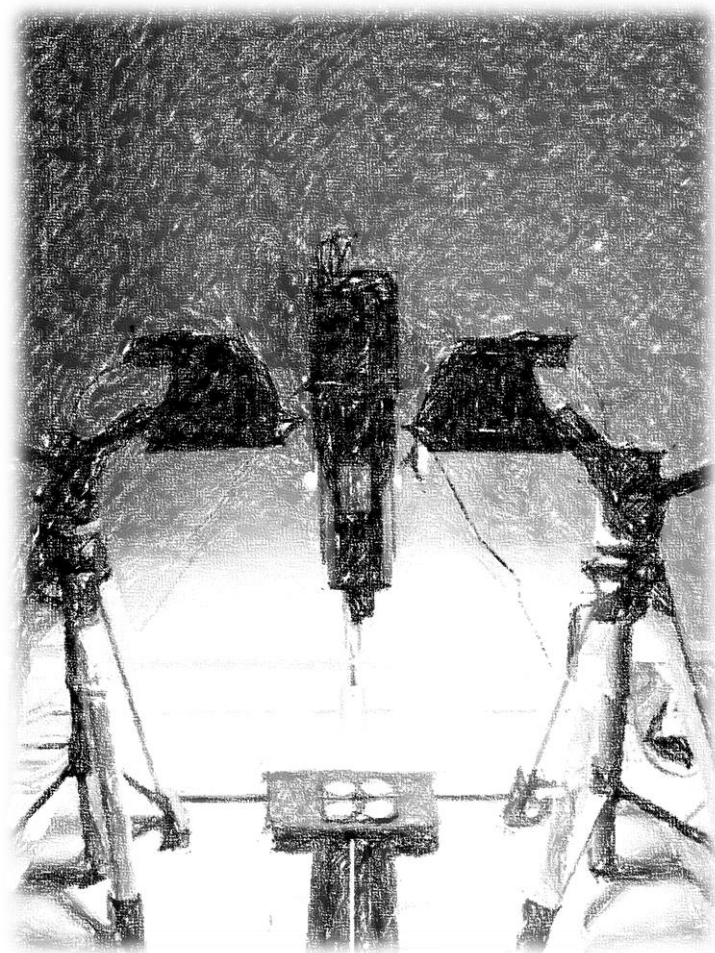
Wester, J. A., T. Kourti and J. F. Macgregor. 1998. Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics* 12(5):301-321.

Worley, B., S. Halouska and R. Powers. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical Biochemistry* 433(2):102-104.

Yang, C.-C., K. Chao, M. S. Kim, D. E. Chan, H. L. Early and M. Bell. 2010. Machine vision system for on-line wholesomeness inspection of poultry carcasses. *Poultry Science* 89(6):1252-1264.

Yang, C.-C., M. S. Kim, P. Millner, K. Chao, B.-K. Cho, C. Mo, H. Lee and D. E. Chan. 2014. Development of multispectral imaging algorithm for detection of frass on mature red tomatoes. *Postharvest Biology and Technology* 93(1):1-8.

5.5. *“Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, Chemometrics and Intelligent Laboratory Systems, 172 (2018) 90-99.*





Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins



Cristóbal Garrido-Novell^{*}, Ana Garrido-Varo, Dolores Pérez-Marín, José Emilio Guerrero

Faculty of Agriculture and Forestry Engineering, Non-Destructive Spectral Sensors Unit, University of Cordoba, Spain

ARTICLE INFO

Keywords:

Protein animal by-products
Near infrared hyperspectral imaging
Pixel selection algorithms
Partial least squares discriminant analysis
Spectral and textural analysis
Classification trees

ABSTRACT

This paper proposes a method based on Near Infrared Hyperspectral Imaging for discriminating between pork, poultry and fish species in processed animal protein meals. First, an investigation was conducted into the possible importance of incorporating into the discrimination models anomalous (or singular) pixels as probable discriminant pixels for each species. Subsequently, partial least squares discriminant analysis (PLS-DA) spectral and textural models were constructed. The former reflected the spectral information (spectral trace), and the latter the spatial (textural trace) information based on different groups of features. Finally, the spectral and textural information was integrated using classification trees, to ascertain whether the combined use of such information represented an improvement in accuracy in the effort to discriminate between species.

The method was applied to a set of 40 pork, 40 poultry and 40 fish meals analysed in the 1000–1700 nm range. Models were then tested using an external validation set comprising 45 samples (15 pork, 15 poultry and 15 fish meals). The results demonstrated that combining spectral and appearance characteristics in a single classification tree generated better classification results for the samples used in the study (92% correct) than when using the PLS-DA spectral model (83% correct).

1. Introduction

Processed Animal Protein (PAP) is an EU term applied to rendered materials belonging to category 3 by-products (fit for human consumption). These can be derived from animal by-products (ABP) or land animal protein (LAP) and from fish. PAPs and LAPs are now highly valued by a number of industries, and have become a major complete-feed ingredient for pets [1].

Article 11 of Regulation (EC) No 1069/2009 of the European Parliament and of the Council of 21 October 2009, prohibits the feeding of terrestrial animals of a given species (e.g. pork, poultry), other than fur animals, with PAP derived from the bodies or parts of bodies of animals of the same species (intra-species recycling). Article 11 also prohibits the feeding of farmed fish with PAP derived from the bodies or parts of bodies of farmed fish of the same species. As stated in its Strategy paper on Transmissible Spongiform Encephalopathies for 2010–2015 [2], the Commission recognizes that a lifting of the ban on the use of PAP from non-ruminants to non-ruminants could be envisaged, provided that the existing prohibition on intra-species recycling is

maintained and only if validated analytical techniques to determine the species origin of PAP are available. Following similar lines, in 2013 the Commission adopted a first review of the feed ban provisions in order to allow aquaculture animals to be fed with PAP derived from non-ruminant farmed animals.

However, despite the huge amount of effort invested, there is still no validated diagnostic method capable of detecting the presence of porcine or poultry material in feed. Therefore, it would not be possible to control the correct implementation of the prohibition on intra-species recycling should the use of PAP of porcine origin in poultry feed or the use of poultry PAPs in pig feed be re-authorised [3]. In response to the regulatory pressure, the industries concerned, via their national and international associations, have made great efforts to collaborate in and finance scientific research [4]. With regard to the lifting of the feed ban on non-ruminant PAP-use within the EU – poultry PAP to pigs and porcine PAP to poultry – the European Feed Processors Rendering Association (EFPPRA) is still waiting for Polymerase Chain Reaction tests that must be available before the EC will consider lifting the ban. The problem is that the tests cannot distinguish between legal ruminant

^{*} Corresponding author.

E-mail address: crisgano8459@hotmail.com (C. Garrido-Novell).

ingredients, such as milk, and ruminant PAP. EFRA has consequently claimed that this can easily be resolved by testing the ingredients – the PAP – rather than the finished feed, where more ingredients are mixed into the final diet.

One of the techniques that has shown the most promising results in distinguishing between animal species in animal-derived feeds is NIR microscopy combined with chemometric techniques. De la Haba et al. [5] were able to distinguish between particles of terrestrial animals by modelling their NIR microscopy data with a Support Vector Machine (SVM). Pérez-Marín et al. [6] were able to distinguish between particles of terrestrial and aquatic animals. In their case, the methodology proposed was based on the prior selection of animal particles and the subsequent application of a K-nearest-neighbours (KNN) model. This methodology showed its potential for detecting the presence of animal content in composite feeds. Other studies, such as that of Boix [7], went further in this direction. The greatest drawbacks implicit in this technique in terms of its industrial application are that it is highly time-consuming, requires specific training and involves the use of carcinogenic substances to extract the fragments under analysis [6]. Conventional NIR technology, overcoming these problems, has been used to determine the percentage of animal content in composite feeds, as well as to distinguish between feeds originating from different species. De la Haba et al. [8] distinguished, with a 90% success rate, not only between ruminant- and non-ruminant-based feeds, but also between poultry-, porcine- and ruminant-based feeds, although it should be pointed out that the error rates are high in terms of the detection limits stipulated by the EU.

One of the preliminary stages in any protocol involving chemometric treatment is discarding anomalous spectra [9]. The application of this to hyperspectral images at the level of pixels, whether or not representative pixels are chosen [10], incurs the risk of eliminating pixels that, despite their low proportional incidence, are exclusive to a particular type of feed and are therefore discriminant. Riccioli et al. [11] proposed a method based on hyperspectral images to distinguish between feeds derived from terrestrial animals and fish using a selection of representative pixels, achieving a rate of precision at the pixel level in excess of 99%. Riccioli et al. [10] compared the performance of classic NIRS spectroscopy with hyperspectral NIRS imagery in the detecting and quantification of ruminant feed in PAPs. The results showed that whereas the hyperspectral image had greater potential in terms of species differentiation, due to the qualitative information it provided, the quantification of contamination with ruminant feed was established with greater accuracy by classic NIRS.

Another method that has been investigated recently focuses on extracting distinctive geometrical features by analysing microscopic imagery. Pinotti et al. [12], tried to distinguish between poultry and mammal particles in animal feeds by extracting geometrical attributes. Although significant differences were detected between classes, it was only possible to make a distinction between averages, not individual particles, owing to the major overlaps between classes. Using the same method, Ottoboni et al. [13] found average differences between porcine and bovine particles, but the method proved unsuitable for differentiating between individual particles. Yao et al. [14] were able to distinguish between particles of fish and other animal meals on the basis of their geometrical properties. Discrimination between poultry and mammals proved to be more difficult however, with a discrimination rate of 93% being achieved. Differentiation between porcine and bovine particles was not possible.

Studies into the more traditional application of multispectral image analysis in foods have focused on spectral signature rather than imaging features. Analysis of the hyperspectral image however enables a simultaneous focus on both spectral and geometrical aspects. Although applications of this geometrical approach have been limited to certain fields, such as the segmentation of aerial or satellite imagery to identify land use, they have also in recent years been identified as a valuable tool in agri-food areas [15].

A general procedure for extracting textural features of images in the spatial domain was presented by Haralick et al. [16]. A co-occurrence matrix (GLCM) is a square matrix with elements corresponding to the relative frequency of occurrence of pairs of grey level of pixels separated by a certain distance in a given direction (0° , 45° , 90° or 135°). The standard procedure extracts the textural attributes of monochrome images. When it is applied to a hyperspectral image, the extraction is usually also carried out using monochrome images: in a specific band or in projections on the principal components. In the realm of food technology, Garrido-Novell et al. [15] applied textural analysis to images of slices of ham projected onto the first principal component for qualitative analysis of the slices with a hyperspectral image. In the realm of remote sensing, studies such as that of Huang et al. [17] propose methods of applying textural analysis to multispectral and hyperspectral images, using alternative methodologies to the analysis of principal components.

This study is aimed at evaluating a methodology based on a model that combines spectral and texture information extracted from hyperspectral NIR imaging to improve accuracy in discrimination between processed pork, poultry and fish proteins.

2. Experimental

2.1. Image acquisition

A total of 120 meal samples belonging to different categories were analysed: 40 pure poultry meal samples, 40 pure swine meal samples and 40 pure fish meal samples. The samples contained particles of scale, hair, feather, blood, grease, skin, muscle, and bone from either fish or terrestrial animals and were obtained from several rendering plants.

One gram (1 g) of each sample was used for analysis. Samples were analysed with a line scan NIR imaging system comprising a camera (XEVA-1.7-320 CCD, Xenics, Leuven, Belgium), a spectrograph (ImSpector V10E, Specim, Oulu, Finland) ranging from 900 to 1700 nm, two 250 W halogen lamps and a control step platform (Velmex, Bloomfield, NY). The image size was 256×200 pixels with a spatial resolution of $0.5 \text{ mm}^2/\text{pixel}$ and a spectral resolution of 3.3 nm.

A 2601-pixel Region of Interest (ROI) was selected in order to avoid those parts of the image not belonging to the sample (Fig. 1).

Raw images were transformed into reflectance images using a 99% reflectance ceramic board. A reflectance reference was obtained by

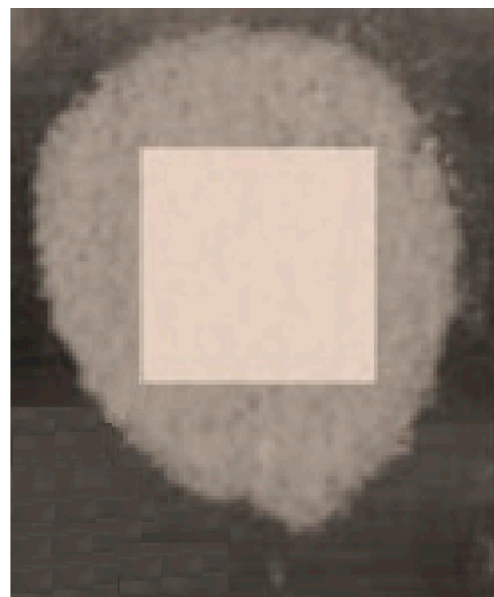


Fig. 1. Sample of meal on a steel plate and ROI.

closing the camera objective. The images were subsequently transformed into absorbance images.

2.2. Discriminant pixels

Two statistics were used in this regard: the Hotelling t^2 and Q-residual (95% confidence level) (Eigenvector Research, Wenatchee, WA). These pixels, obtained out of the control limits for both parameters, were extracted and joined to pixels of other species in a Principal Component Analysis in order to compare different spectral behaviours. The number of latent variables was selected looking for the minimum cross validation error.

Baseline drifts due to light scattering affect the quality of raw measurements. Therefore, good spectral pre-processing is mandatory. In this study, the spectra were pre-treated with a smoothing, a Standard Normal Variate (SNV) and a Mean Centre (MC). The SNV approach effectively removes the multiplicative interferences of scatter [18]. Mean centring translates the axes of the coordinate system to the centre of gravity of the data. After mean centring, each row of the mean-centred data explains only how that row differs from the average sample in the original data matrix. Mean centring avoids detecting the differences among the mean values of different variables as significant directions of variability; scaling to unit variance makes the analysis independent of the measurement units, thus enabling the simultaneous analysis of variables with values of very different magnitudes [19].

2.2.1. Partial least squares discriminant analysis (PLS-DA)

PLS-DA models were developed by assigning a reference value (dummy variable) to each sample. A sample was considered to be correctly categorized if the predicted value was above a threshold, which was calculated using the Bayesian threshold calculation [20]. As PLS-DA models yield a prediction for each sample and each class, when comparing between models, the probability of each sample of belonging to each class was used. Probability was calculated according to the method describe by N. F. Pérez et al. [21].

The average spectrum of each sample was calculated. PLS-DA was performed and evaluated using the statistics' sensitivity, specificity and classification error. The sensitivity of a class model is defined as the proportion of objects belonging to that class correctly identified by the mathematical model, while specificity refers to the proportion of objects not belonging to the class, classified as foreign [22]. This first model summarised the spectral information. A second PLS-DA model was built using curves of texture in order to summarise textural and thus spatial information. PLS-DA models were built using PLS-toolbox from Matlab 7.6.8. The 120 samples were randomly divided into 96 calibration samples and 24 validation samples. Both models were integrated into just one model using classification trees developed using the Rattle package [23], for R 3.3.2. Predictions and classification were tried as input variables to build these.

Standard Normal Variate (SNV) and Mean Centre (MC) were applied to average spectra. Contiguous Block Cross validation was used. The number of latent variables was selected looking for the minimum cross validation error.

In this study, the textural features were calculated from the Co-occurrence Matrix [16], averaging the results obtained in all possible directions and for distances equal to 1, 3 and 5 pixels. The following four GLCM parameters were calculated using Matlab 7.6.8 to express texture: contrast, homogeneity, energy and correlation.

$$Contrast = \sum_{i,j} |i - j|^2 p(i,j)$$

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1 + (i - j)^2}$$

$$Energy = \sum_{i,j} p(i,j)^2$$

$$Correlation = \sum_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] p(i,j)$$

... where i and j are pixel reflectances and p_{ij} is the frequency of co-occurrence of i and j reflectances.

Each textural parameter expresses how pixel intensities are distributed throughout the image. The 212 textural parameters corresponding to the 212 wavelengths were joined in order to build a curve. This curve is the spatial counterpart of the spectral print. Each NIRS spectrum summarizes the spectral information of a sample by a curve of absorbances or reflectances. This spectrum is usually called spectral print. The contrast texture parameter summarizes how this absorbance is distributed in each monochrome image. From the representation of these values in a graph in which the “x” axes is the wavelengths one (each monochrome image at each wavelength) and the “y” axes is the average “Contrast” parameter for each monochrome image, arises this curve that we have called contrast curve, in the same way as the NIRS spectrum of a sample is called its spectral print. This curve summarizes the textural information of the entire hypercube.

Textural curves were pre-treated using smoothing, SNV, first derivative and Mean Centre. The 15 extreme wavelengths were discarded. Contiguous Block Cross Validation was used. The number of latent variables was selected looking for the minimum cross validation error.

2.2.2. Classification trees

Classification and regression trees are machine-learning methods for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree [24]. Pruning rules change from one algorithm to another. The pruning rules used by the “rpart” package, in Rattle, are minimum number of samples per node and minimum number of samples resulting from splitting and complexity, which express the maximum allowable development [23].

The model was calibrated with the same 96 samples used for calibrating both the spectral and spatial PLS-DAs. It was validated with the same 24 samples used in previous models. The sample scores and the predictions from the PLS-DA models were tried as input variables. This prediction algebraically combines the scores to obtain a numerical prediction. The following steps and pruning rules were followed in order to build the tree:

1. The minimum number of samples that a node has in order to be subdivided (9 and 6) and the minimum number that must result from this partition (3 and 2) were fixed. The tree was developed up to the maximum of complexity (0) using the Gini index.
2. The number of partitions for which the relative cross-validation error is minimal were selected, provided that the relative cross-validation error was not too far from relative error of calibration. From this point, the increase of the complexity implies an increase of the artificiality.
3. The tree was rebuilt by limiting its development (pruning) to the above-mentioned partition level.

3. Results and discussion

3.1. Anomalous pixels

By building multivariate models based on Principal Components or Latent Variables, the samples are being projected onto planes on which the covariance – which is by definition an average parameter – is

maximal, such that the existence of pixels capable of providing useful information is being overlooked. This effect is greater when we discard, as a preliminary step to chemometric analysis, anomalous pixels of samples consisting in turn of 2601 0.5 mm² sub-samples. Locating singular pixels or groups of pixels, and analysing them independently from the rest of the pixels, could enable both the differentiation of samples and the identification of characteristic groupings of pixels.

The anomalous pixels exhibited the characteristics shown in Figs. 2 and 3. Fig. 2 reflects the anomalies of four poultry (A), four fish (E) and four porcine (P) hyperspectral images (2601 pixels each). Pixels were chosen with a high Hotelling t^2 and a high Q-residual (95% confidence), in other words, pixels distant from the centre of the PCA model and with little weight in it.

Fig. 3 shows the anomalous pixels of four images representative of each species. Only the anomalous pixels of poultry (a) and swine (p) exhibit differences from the bulk of the pixels when they are shown together with the pixels from an image of fish meal (E). By contrast, the anomalous fish pixels are not located outside the area formed by the non-anomalous pixels of poultry (A) and swine (P). These findings contrast with those obtained by Riccioli et al. [11] who observed greater amplitude in the fish meal histogram compared to the histogram for terrestrial meals in the fat-absorption band (1190 nm). By contrast, for the complete spectral range of 900–1700 nm, the anomalous poultry and porcine pixels were found to be more heterogeneous than those of fish.

These relative distributions provide no information about the existence of singular pixels, given that the anomalous pixels are grouped in the same way as the non-anomalous pixels with respect to those of other species. This is evident from the representation of the non-anomalous

pixels (that is, the bulk of the four samples remaining pixels after extracting the anomalous ones) by pairs of species (Fig. 4).

Pinotti et al. [12,25] gave a geometrical characterisation of the densest sedimented particles (bones, etc.) from poultry, mammal and fish meals. They observed an average area for the poultry particles of 65.76 μm^2 , with a greatest dimension of 17.94 μm ; 99.72 μm^2 and 19.23 μm for mammals; and 100 μm^2 and 35 μm for fish.

The absence of singular pixels could be explained either by the size of the singular particles compared to the size of each pixel, or by the scarcity of these particles in each of the pixels, causing their spectral signals to become diluted. The area covered by each pixel, 0.5 mm², is 5000 times greater than the largest particles observed by Pinotti et al. [12,25], the fish particles of around 100 μm^2 .

3.2. Classification models

3.2.1. Spectral classification model

Various PLS-DA models were developed with the pre-treated average spectra, using different pre-treatments of the spectral data. Neither the calibration set nor the cross-validation set was modified while extracting anomalous samples, with the aim of being able to make later comparisons with other classification models. Indeed, the distribution of samples between calibration and validation used for this model was that used by the rest of the classification models. The model that exhibited the least classification error in cross validation was chosen out of all the models obtained.

As is evident from Table 1, the best-classified group is that of fish meal, with an error rate of 0% both in cross validation and in external

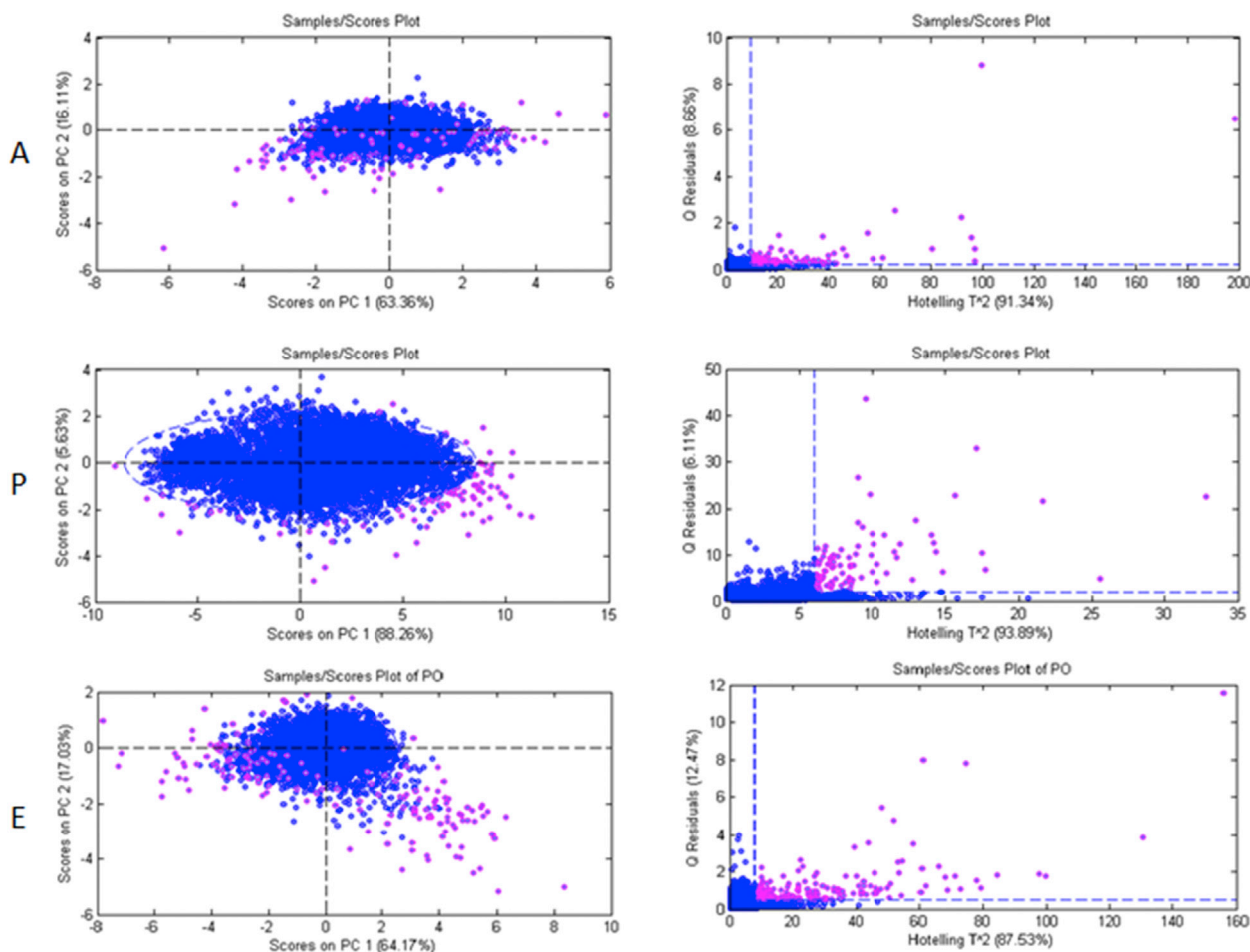


Fig. 2. Left: PC1 vs. PC2 of (A) poultry meal pixels; (P) pork meal pixels; (E) fish meal pixels. Right: Hotelling's T^2 versus Q-residual plot of (A) poultry meal pixels (from a 4 Principal Components model); (P) pork meal pixels (from a 4 Principal Components model); (E) fish meal pixels (from a 2 Principal Components model).

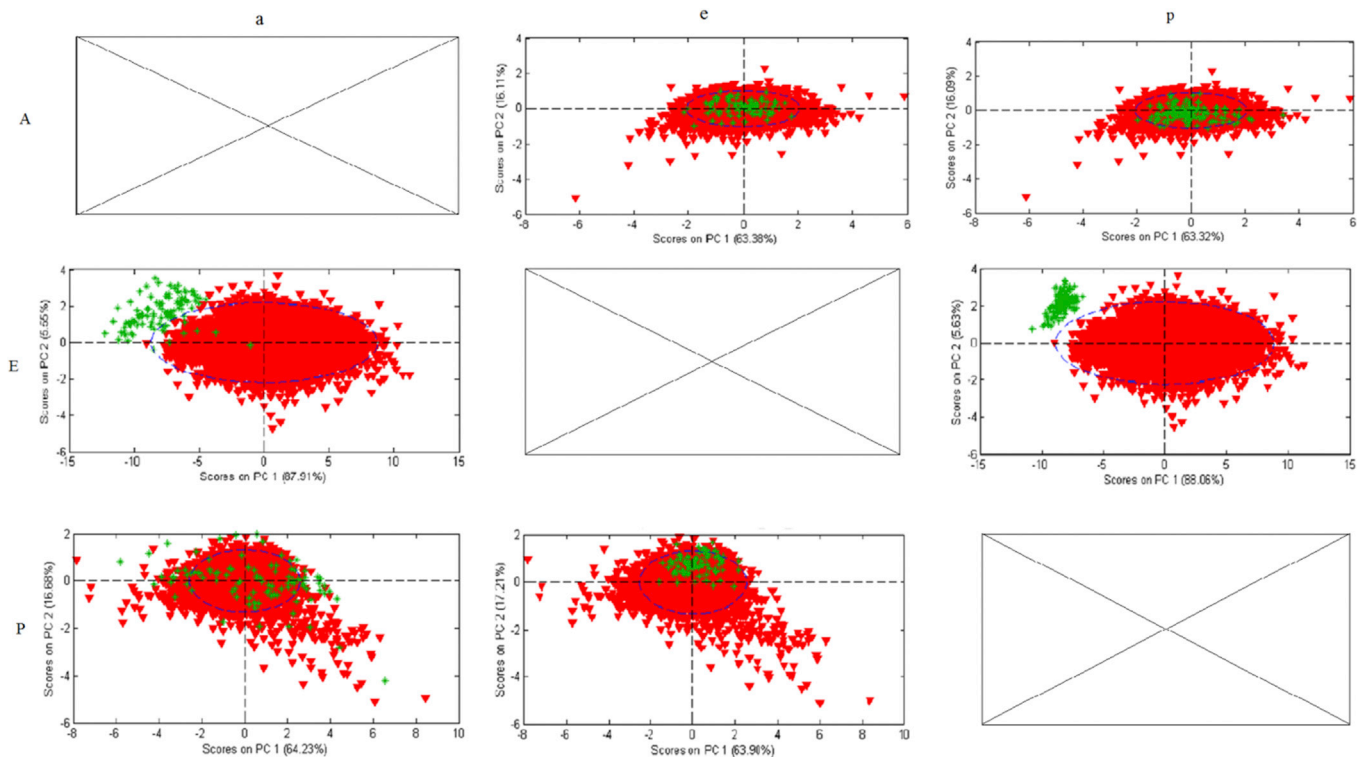


Fig. 3. Anomalous poultry (a), fish (e) and porcine (p) pixels, in green, on poultry (A), fish (E) and porcine (P) pixels, in red. All are shown on the two first principal components. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

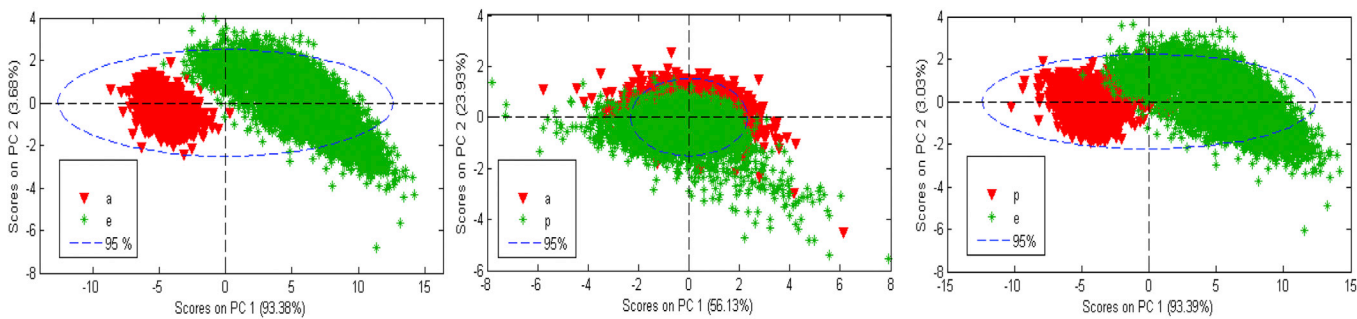


Fig. 4. Projections on the two first principal components of the pixels of the four poultry (a), fish (e) and porcine (p) samples (the same samples used in Fig. 1) together with one of the other two species.

Table 1
Results of the classification using spectral PLS-DA in cross validation (CV) and external validation (EV). Percentage of correctly classified samples per species.

		Predicted					
		CV			EV		
		Poultry	Fish	Pork	Poultry	Fish	Pork
Actual	Poultry	29	0	3	6	0	2
	Fish	0	32	0	0	8	0
	Pork	2	0	30	2	0	6
	Hits (%)	91	100	94	75	100	75

validation. The classification errors for poultry and porcine samples in cross validation were 9% and 6% respectively, although in external validation they rose to 25% in both cases. These samples could have been included in the calibration set or they could even have been deemed anomalies. The challenge of the methodology being proposed is to reduce these rates of error by taking the spatial information into consideration.

The 83% overall success rate achieved by this model is lower than that obtained by De la Haba et al. [5], using classic NIRS, who managed to obtain, also using a PLS-DA model, an overall success rate of 90%, differentiating between poultry meal, porcine meal and mixtures of ruminants with other species for a group of 352 samples. Other studies, for example Riccioli et al. [11], despite using more similar groups, both in terms of size and classes, are not comparable owing to the fact that they address the building of models from another perspective: they apply their models to pixels rather than complete samples, obtaining success rates close to 100% for all classes.

3.2.2. Textural classification model

Despite the homogenisation that animal by-products are subjected to when they are processed, the samples, and therefore their images, are not homogeneous. The absence of discriminant pixels does not in itself prevent the possibility of there existing characteristic distributions of pixels. It is possible for there to be characteristic distributions linked to textures that are common to all species, and are therefore not

singular in themselves, but acquire distinct patterns depending on the species.

Assuming such characteristic distributions of pixel intensities exist, applying the textural parameters to all the bands of the hyperspectral image combined with a multivariate analysis of the resulting curves may prove to be a sufficiently exhaustive method for identifying them. The method being proposed pursues the same aim as those put forward by Huang et al. [17], using clustering and sparse representation algorithms.

Fig. 5 shows the mean contrast curves for pixels separated by another two pixels. This reveals that the contrast values vary according to the wavelength in question, and that the contrast curves vary between species. The curves derived from the other two separations between pixels (contiguous pixels and four pixels) show a similar appearance in terms of shape, but distinct from the above in terms of their displacement along the Y axis.

Tables from 3 to 5 show confusion matrices for the cross validation and external validation of the contrast curve models. It is evident that although the models based exclusively on textural information discriminate to an acceptable degree, they are not sufficiently informative. It may be thought that this is due to specific characteristics of the samples, unconnected to their proteinic composition, such as finer grades of milling, different moisture levels, etc. The intention underlying the pre-treatment to which the images were subjected prior to being processed was to eliminate the effects of light dispersal associated with the varying sizes of the particles. Even so, the only way of verifying the validity of these models is applying them to groups of feeds subjected to different types of processing. Such an experiment lies outside the scope of the present study.

The best models were obtained using information about pixels separated by another four pixels, i.e. non-contiguous pixels. This means that the most representative differences as far as the contrast parameter between species is concerned emerges between pixels separated by another four pixels. One explanation among many could be that the size of the groupings enabling differentiation between species exceeds that of the pixel. Thus, the greater frequency of pairs of high-contrast reflectance values occurs between this grouping and the pixels that lie beyond its domain (beyond the four pixels). The differences between species could be due to distinct pairs of intensity levels for distinct species between the grouping and its surroundings or the different sizes of the groupings between species.

If the results shown in Table 1 are compared with those of Tables 2–4 it becomes apparent that despite providing discriminatory information, the classification errors of the textural models exceed those obtained with the spectral classification models.

Table 2

Classification results using PLS-DA for contrast curves (contiguous pixels) in cross validation (CV) and external validation (EV).

		Predicted					
		CV			EV		
		Poultry	Fish	Pork	Poultry	Fish	Pork
Actual	Poultry	25	1	6	5	1	2
	Fish	2	29	1	0	8	0
	Pork	8	5	19	0	0	8
	Hits (%)	78	91	59	63	100	100

Table 3

Classification results using PLS-DA for contrast curves (pixels separated by two pixels) in cross validation (CV) and external validation (EV).

		Predicted					
		CV			EV		
		Poultry	Fish	Pork	Poultry	Fish	Pork
Actual	Poultry	27	1	4	6	2	0
	Fish	0	29	3	0	7	1
	Pork	9	2	21	1	0	7
	Hits	84	91	65	75	88	88

Table 4

Classification results using PLS-DA for contrast curves (pixels separated by four pixels) in cross validation (CV) and external validation (EV).

		Predicted					
		CV			EV		
		Poultry	Fish	Pork	Poultry	Fish	Pork
Actual	Poultry	25	1	6	5	1	2
	Fish	1	30	1	0	8	0
	Pork	8	2	22	0	0	8
	Hits (%)	78	94	69	63	100	100

3.2.3. Classification trees

The spectral and textural models were integrated into one another using classification trees, to ascertain whether the combination of the spectral and textural information would bring about an improvement of the performance seen in the spectral models.

Tables 5 and 6 show the sensitivities and specificities of the models in tree-form. The former shows the proportion of correctly classified positives, while the latter shows the correctly classified negatives. Best models are highlighted in bold.

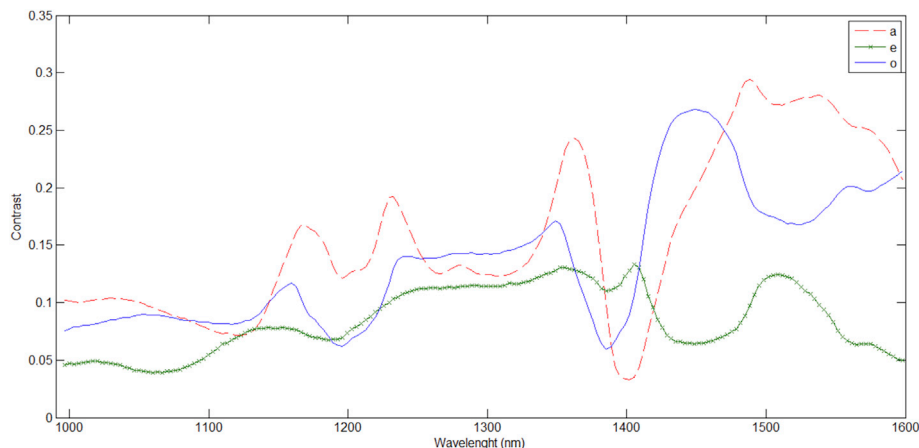


Fig. 5. Mean contrast curves by species (a, poultry; e, fish; o, swine) for pixels separated by another two.

Table 5

Summary of the classification tree constructed on the basis of the quantitative predictions of the PLS-DA models.

		Cross Validation		External Validation	
		Sensitivity	Specificity	Sensitivity	Specificity
Poultry	Contiguous	0.86	1.00	0.50	1.00
	2 pixel separation	1.00	0.94	0.89	0.94
	4 pixel separation	0.86	1.00	0.50	1.00
	All of them	1.00	0.94	0.89	1.00
Fish	Contiguous	1.00	1.00	1.00	1.00
	2 pixel separation	1.00	1.00	1.00	1.00
	4 pixel separation	1.00	1.00	1.00	1.00
	All of them	1.00	1.00	1.00	1.00
Pork	Contiguous	1.00	0.94	1.00	0.75
	2 pixel separation	0.88	1.00	0.88	0.94
	4 pixel separation	1.00	0.94	1.00	0.75
	All of them	0.88	1.00	0.88	0.94

Table 6

Summary of the classification tree constructed on the basis of the latent variables of the PLS-DA models.

		Cross Validation		External Validation	
		Sensitivity	Specificity	Sensitivity	Specificity
Poultry	Contiguous	1.00	0.82	0.75	0.75
	2 pixel separation	1.00	0.88	0.75	0.75
	4 pixel separation	1.00	0.88	0.75	0.75
	All of them	1.00	0.82	0.75	0.75
Fish	Contiguous	1.00	1.00	1.00	1.00
	2 pixel separation	1.00	1.00	1.00	1.00
	4 pixel separation	1.00	1.00	1.00	1.00
	All of them	1.00	1.00	1.00	1.00
Pork	Contiguous	0.63	1.00	0.50	0.88
	2 pixel separation	0.75	1.00	0.50	0.88
	4 pixel separation	0.75	1.00	0.50	0.88
	All of them	0.63	1.00	0.50	0.88

The classification trees shown in Table 5 were constructed on the basis of the quantitative predictions of the spectral and textural PLS-DA models. Six variables – three spectral and three textural – were included in their construction. Each of these is identified in the tables by the textural model used, the spectral model being common to all of them: whereas four types of textural model were developed (contrast of contiguous pixels, pixels separated by another two, pixels separated by four pixels; and lastly a model featuring all the preceding variables). The number of variables therefore ranged from six, for the simple trees, to 12 for the complete tree.

The data shown in Table 6, by contrast, were obtained by projecting the spectral and textural traces onto the latent variables (“scores”) of the spectral and textural PLS-DA models. Depending on the model, more or fewer latent variables were incorporated into the tree. The four latent variables of the spectral models were common to all trees. The textural models on the other hand contributed to the tree with three, seven and three latent variables for the models of contrast curves between contiguous pixels, separated by two pixels and separated by four pixels, respectively. The number of variables therefore ranged between seven for the simple trees and 17 for the complete tree.

Comparing first the trees constructed on the basis of scores with those constructed on the basis of quantitative predictions (dummy variables), it is evident that the latter are much more sensitive and specific. Despite the fact that the scores gather more information than the predictions (which are an algebraic combination of the former), they do so in a cruder fashion. Contrary to what was assumed at the outset, the stepped system of decisions seems to fit models based on quantitative predictions better.

In both types of model (based on the scores or on the predictions, of Tables 5 and 6), the tree that exhibited the greatest sensitivities and specificities both in terms of external validation and cross validation was the one that included the textural model for pixels separated by another two pixels. The models with all the latent variables (“All”) made no improvement upon the results of the simplest tree models.

Comparing the parameters derived from the best tree model (Table 5, in bold), which combines the single spectral model with the model based on quantitative predictions of the textural model, with those obtained from the spectral PLS-DA model (Table 7), the greater robustness of the tree model becomes evident, for both sensitivity and specificity in the three species in cross validation and external validation. The only exception was the sensitivity of the porcine sample in cross validation, where the tree model obtained 0.88 as opposed to the 0.94 of the spectral PLS-DA model.

In absolute terms, the greater goodness of fit of the tree model can be appreciated by comparing the confusion matrices (Table 8, showing the best tree model, with those of the spectral PLS-DA model, displayed in Table 1). Specifically, the porcine and poultry samples that were incorrectly classified in external validation were reduced by half in both cases: from two to one. In studies such as that of De la Haba et al. (2007b), carried out using classic NIRS, a 90% correct classification in external validation was achieved for poultry, porcine and bovine samples; including the textural variables the authors of the present study attained 92%, while with the spectral model alone a correct classification rate in external validation of 83% was achieved (Table 1). With this model, the present authors exceeded the overall correct percentages obtained by De la Haba et al. [5], using classic NIRS, differentiating between poultry meal, porcine meal and ruminant meal mixed with other species for a group of 352 samples.

Fig. 6 shows the classification tree that obtained the best results. It is apparent how the textural variable ($P_{C_3} \geq 0.65$) serves to obtain a node of porcine meal samples of 100% purity (node 11). In external validation, five porcine samples out of a total of eight are classified in this node correctly. The other gradient ($P_{C_3} < 0.65$) provides the location of the two classification errors in external validation. Three poultry and one porcine sample were correctly classified in the former, whereas in the latter one poultry sample was incorrectly classified and two porcine samples were correctly classified. Although the incorporation of the textural variable into the tree enhanced the output of the model, there continues to be a degree of unresolved indeterminism linked to these two samples, one poultry and one porcine.

To determine which bands had most influence on the best of the textural models the “Variable Importance in Projection” (VIP) parameter (Fig. 7) was used; this estimates the importance of each variable used in the PLS model. A VIP score close to or greater than one indicates the importance of a band in the construction of the PLS-DA model.

The wavelengths that obtained a VIP score greater than one were those of 1192.6 nm (band 60), 1215.9 nm (band 67), 1245.8 nm (band

Table 7

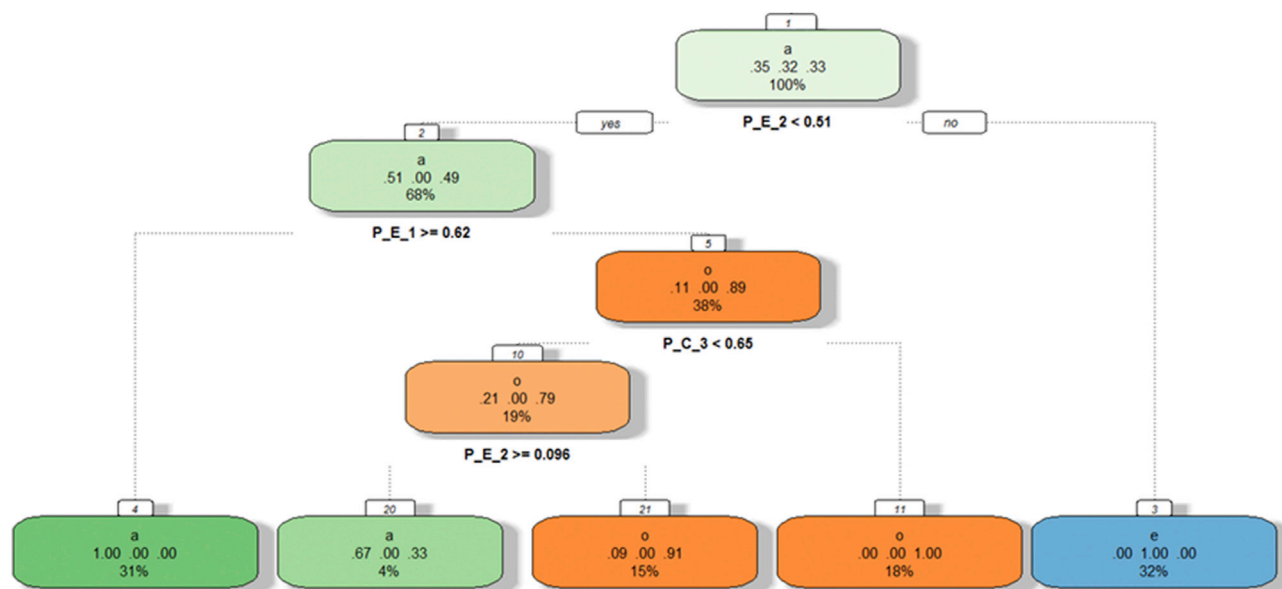
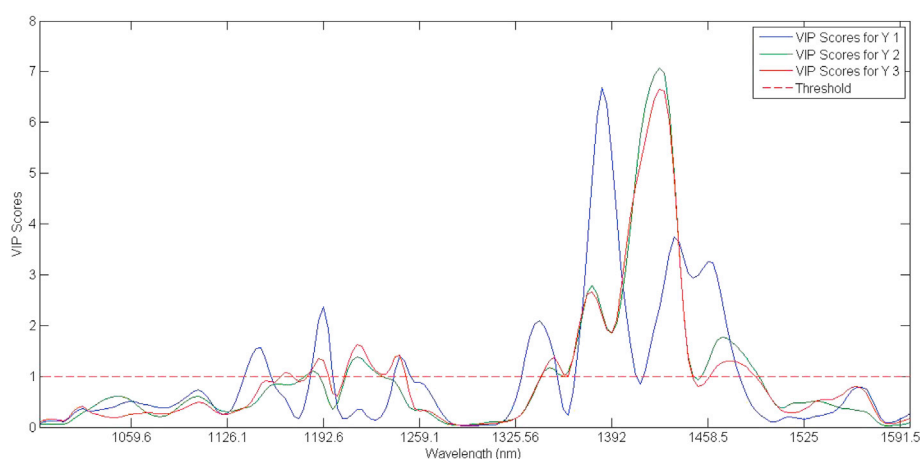
Summary of the classification using spectral PLS-DA in terms of specificity and sensitivity.

		Cross Validation		External Validation	
		Sensitivity	Specificity	Sensitivity	Specificity
Poultry		0.90	0.97	0.75	0.88
Fish		1.00	1.00	1.00	1.00
Pork		0.94	0.93	0.75	0.88

Table 8

Results of the tree classification based on the predictions of the PLS-DA (calibration), CV (cross validation) and EV (external validation).

		Predicted								
		Calibration			CV			EV		
		Poultry	Fish	Pork	Poultry	Fish	Pork	Poultry	Fish	Pork
Actual	Poultry	24	0	1	7	0	0	7	0	1
	Fish	0	23	0	0	9	0	0	8	0
	Pork	1	0	23	1	0	7	1	0	7
	Hits (%)	96	100	96	100	100	88	88	100	88

**Fig. 6.** Classification tree (spectral PLS-DA + textural PLS-DA) that obtained the best results. Into the boxes the percentage of sample from the whole set and the proportion of each species for each node can be seen.**Fig. 7.** VIP scores for the textural model included in the best tree model.

76); 1342.2 nm (band 105), 1385.4 nm (band 118), 1425.3 nm (band 130), 1435.3 nm (band 130) and 1458.5 nm (band 140).

The first group of peaks is located around the zone corresponding to fat (1200 nm) while the second is grouped around the water zone (1450 nm). Lastly, the 1458.5 nm wavelength is located close to the zone corresponding to the amino acids (1500 nm). The two peaks of greatest importance are those corresponding to the 1385.4 nm and 1425.3 nm wavelengths [25].

These data show that it was the importance of the distribution of the water in the samples, captured by the textural curves, that enabled the classification model to be improved. Venturing into the realm of speculation, and pending the validation of the model with larger and more varied sets capable of showing that this phenomenon is not due to the individual characteristics of the samples used, the data obtained appear to indicate that the water is distributed in a distinct way depending on the species.

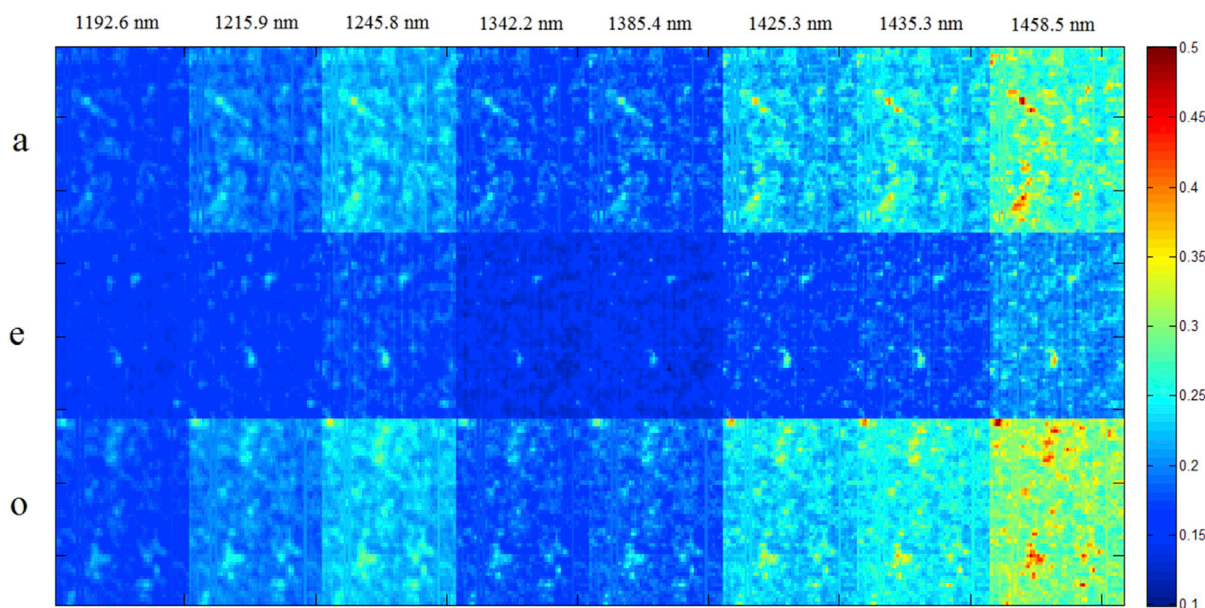


Fig. 8. Bands in which the VIP score was greater than one.

At first glance, the contrast differences between pixels are imperceptible (Fig. 8), making textural analysis indispensable.

4. Conclusion

The analysis of anomalous pixels did not reveal the existence of singular pixels in their midst that, separating them from the behaviour of the majority, would enable discrimination between species. Their behaviour matched that of the non-anomalous pixels. One of the possible causes of this absence is the excessive size of each pixel relative to the size of the particles, leaving open as a future line of research the study of spectral mixtures in such pixels based on the spectral signals collected through NIRS microscopy, as well as the adaptation of the NIRS imaging equipment to achieve higher spatial resolutions.

A methodology involving analysis of the textural characteristics of hyperspectral images was put forward. The use of regression trees enabled the spectral and textural information to be integrated into a single model, so that better classification results than those of the spectral model could be obtained from the samples used (92% versus 83%). The tree model improved the classification by species obtained by such authors as De la Haba et al. [5]. A 92% correct classification in external validation was achieved for poultry, porcine and fish samples using textural variables and trees versus a 90% attained by De la Haba et al. [5].

For these particular samples, the distribution of the water in the samples had a bearing upon improving the classification model. Future studies will need to corroborate this hypothesis using larger and more varied sets of samples, and to evaluate simpler models based on specific bands or ratios of bands that would enable the processing of such voluminous data sets as hyperspectral images to be speeded up.

Conflict of interest

The authors declare they do not have any conflict of interest in connection to the present paper.

Acknowledgements

This study was carried out within the framework of the *Excellence Project AGR-5129 “MEMS near infrared spectral sensors and hyperspectral imaging for in situ and non-destructive analysis of foods and feeds”*, financed by the Andalusia Regional Government.

References

- [1] A. Garrido-Varo, S. Vega, F. Maroto-Molina, M.J. De la Haba, D. Pérez-Marín, On-site quality control of processed land animal proteins using a portable micro-electro-mechanical-systems near infrared spectrometer, *J. NIRS* 24 (1) (2016) 47–58.
- [2] European Commission, Transmissible Spongiform Encephalopathies Roadmap 2, Communication to the European Parliament and the Council, 2010.
- [3] European Commission, Regulation (EU) N° 56/2013, Amending Annexes I and IV to Regulation (EC) N° 999/2001 of the European Parliament and of the Council Laying Down Rules for the Prevention, Control and Eradication of Certain Transmissible Spongiform Encephalopathies, 2013.
- [4] S. Nates, Future research for the rendering industry, in: D.L. Meeker (Ed.), *Essential Rendering: All About the Animal By-Products Industry*, National Renderers Association, 2006, pp. 274–276.
- [5] M.J. De la Haba, A. Garrido-Varo, D.C. Pérez-Marín, J.E. Guerrero, Near infrared analysis as a first-line screening technique for identifying animal species in rendered animal by-product meals, *J. Near Infrared Spectrosc.* 15 (4) (2007b) 237.
- [6] C. Pérez-Marín, T. Fearn, J.E. Guerrero, A. Garrido-Varo, A methodology based on NIR microscopy for the detection of animal protein by-products, *Talanta* 80 (2009) 48–53.
- [7] A. Boix, J.A. Fernández Pierna, C. von Holst, V. Baeten, Validation of a near infrared microscopy method for the detection of animal products in feeding stuffs: results of a collaborative study, *Food Addit. Contam. Part A* 29 (2012) 1872–1880.
- [8] M.J. De la Haba, J.A. Fernández Pierna, O. Fumière, A. Garrido-Varo, J.E. Guerrero, D.C. Pérez-Marín, Discrimination of fish bones from other animal bones in the sedimented fraction of compound feeds by near infrared microscopy (NIRM), *J. Near Infrared Spectrosc.* 15 (2007a) 81–88.
- [9] G.C. Marten, J.S. Shenk, F.E. Barton, *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*, U.S. Agriculture Handbook No. 643 (Revised with Supplements), Department of Agriculture, 1989, p. 110.
- [10] C. Riccioli, D. Pérez-Marín, J.E. Guerrero-Ginel, T. Fearn, A. Garrido-Varo, Detection and quantification of ruminant meal in processed animal proteins: a comparative study of near infrared spectroscopy and near infrared chemical imaging, *J. Near Infrared Spectrosc.* 20 (6) (2012) 623–633.
- [11] C. Riccioli, D. Pérez-Marín, J.E. Guerrero-Ginel, S. Wouter, A. Garrido-Varo, Pixel selection for near-infrared chemical imaging (NIR-CI) discrimination between fish and terrestrial animal species in animal protein by-products meals, *Appl. Spectrosc.* 65 (7) (2011) 771–781.
- [12] L. Pinotti, T. Fearn, S. Gulalp, A. Campagnoli, M. Ottoboni, A. Baldi, F. Cheli, G. Savoini, V. Dell’Orto, Computer image analysis: an additional tool for the identification of processed poultry and mammal protein containing bones, *Food Addit. Contam. Part A* 30 (10) (2013) 1745–1751.
- [13] M. Ottoboni, F. Cheli, G. Amato, D. Marchis, B. Brusa, M.C. Abete, L. Pinotti, Microscopy and image analysis based approaches for the species-specific identification of bovine and swine bone containing material, *Ital. J. Anim. Sci.* 13 (2) (2014) 377–381.
- [14] Y. Yao, L. Han, Z. Yang, X. Liu, Discriminant analysis of different species of meat and bone meal base on microscopic image processing, *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* 47 (2016), 347–352 and 358.
- [15] C. Garrido-Novell, A. Garrido-Varo, D.C. Pérez-Marín, J.E. Guerrero-Ginel, M. Kim, Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging, *J. Food. Eng.* 153 (2015) 117–123.

- [16] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man. Cybern.* 6 (1973) 610–621.
- [17] X. Huang, L. Xiaobo, L. Zhang, A multichannel gray level Co-occurrence matrix for multi/hyperspectral image texture representation, *Remote Sens.* 6 (2014) 8424–8445.
- [18] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.* 43 (1989) 772.
- [19] M. Wise, N.B. Gallagher, The process chemometrics approach to process monitoring and fault detection, *J. Process Control* 6 (1996) 329–348.
- [20] J.A. Pierna, V. Baeten, A.M. Renier, R.A. Cogdill, P. Dardenne, Combination of support vector machines (SVM) and near infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds, *Chemometrics* 18 (2004) 341–349.
- [21] N.F. Pérez, J. Ferré, R. Boqué, Calculation of the reliability of classification in discriminant partial least-squares binary classification, *Chemom. Intell. Lab.* 95 (2009) 122–128.
- [22] M.P. Derde, L. Kaufman, D.L. Massart, A non-parametric class modeling technique, *Chemom.* 3 (1989) 375.
- [23] G. Williams, Rattle: a data mining GUI for R, *R J.* 1 (2) (2009) 45–55.
- [24] J.R. Quinlan, Induction of decision trees, *Mach. Learn.* 1 (1) (1986) 81–106.
- [25] L. Pinotti, M. Ottoboni, V. Caprarulo, C. Giromini, D. Gottardo, F. Cheli, T. Fearn, A. Baldi, Microscopy in combination with image analysis for characterization of fishmeal material in aquafeed, *Anim. Feed Sci.* 215 (2016) 156–164.

6. CONCLUSIONES.

1. Los mejores valores de SNR y Nitidez Espectral que presentó el equipo modular de barrido lineal usado en la presente tesis doctoral con respecto al equipo de barrido espectral comercial mostraron su plena validez [Capítulo 4.4 de la presente tesis].

2. Los modelos RGB desarrollados pueden cuantificar la evolución del color durante la vida útil de las manzanas. Esta evolución puede modelizarse mediante un modelo cinético de orden cero [Esta conclusión queda reflejada en la publicación: “Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288)]

3. Los resultados demostraron el mayor potencial de la Imagen Hiperespectral con respecto a la Imagen RGB en la explicación de la evolución del color en manzanas, incluso en rangos espectrales supuestamente desfavorables (900 – 1700 nm). Esto permitiría el uso de un único equipo con fines de medición del color y de la composición en manzana. Estas conclusiones conducen a la investigación futura con conjuntos de datos mayores, más variables, diferentes variedades, mayores periodos de almacenamiento, fruta en árbol, etc. [Esta conclusión queda reflejada en la publicación: “Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras”, *Journal of Food Engineering* 113 (2012) 281-288)]

4. Usando el modelo de validación específico desarrollado por Burger y Geladi (2006) se obtuvieron modelos de predicción de precisión moderada y buena para sal y humedad, respectivamente, en cortes transversales de jamón curado [Esta conclusión queda reflejada en la publicación: “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”, *Journal of Food Engineering* 153 (2015) 117-123)].

5. Se mostró la utilidad del análisis de los histogramas de frecuencias y el análisis textural para la caracterización de la distribución de la sal y la humedad en mapas de predicción en cortes transversales de jamón curado [Esta conclusión queda reflejada en la publicación: “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured

ham slices using NIR hyperspectral imaging”, *Journal of Food Engineering* 153 (2015) 117-123)].

6. El parámetro “Correlación” mostró que el gradiente local perpendicular a la superficie del jamón en contacto con la atmósfera se incrementaba a medida que la humedad lo hacía. Los resultados para los parámetros “Homogeneidad” y “Contraste” indicaron que la homogeneidad local de la sal y la humedad no estaba relacionada con el contenido graso, y no era proporcional a la profundidad en el músculo con respecto a la superficie. Ésta era mayor en el músculo *semitendinosus* debido a su posición intermedia entre la superficie en contacto con la atmósfera y la capa de grasa subcutánea [Esta conclusión queda reflejada en la publicación: “Quantification and spatial characterization of moisture and NaCl content of Iberian dry-cured ham slices using NIR hyperspectral imaging”, *Journal of Food Engineering* 153 (2015) 117-123)].

7. Las imágenes de nivel de grasa basadas en FCM y la matriz de coocurrencia aplicadas a la Imagen Hiperespectral son herramientas prometedoras para predecir el nivel de grasa y su distribución en caña de lomo de jamón ibérica [Esta conclusión queda reflejada en la comunicación “Methodology for marbling characterisation of pure Iberian cured loin based on hyperspectral imaging”. *Comunicación oral 16th International Conference on Near Infrared Spectroscopy*. 2013.”]

8. Se obtuvieron modelos basados en PCA y PLSR para detectar contaminaciones en harinas de pescado con harinas de rumiantes. Se alcanzaron tasas de clasificación del 97,9% para rumiantes y 99,4% para harinas de pescado usando un modelo PLSR. Las tasas de píxeles correctamente clasificados fueron de 85,1% para las harinas de rumiante y del 88, 2% para harina de pescado usando un modelo basado en PCA [Esta conclusión queda reflejada en la publicación: “Differentiation of beef and fish meals in animal feeds using chemometric analytic models”, *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)].

9. Las altas tasas de clasificación obtenidas indican el potencial de aplicación de las cámaras hiperespectrales en línea en combinación con modelos quimiométricos para la detección de contaminaciones con harinas de rumiantes en harinas de pescado [Esta conclusión queda

reflejada en la publicación: “Differentiation of beef and fish meals in animal feeds using chemometric analytic models”, *Journal of Biosystems Engineering*, 40(2):153-158. (2015.6)].

10. En la diferenciación entre harinas de distintas especies, el análisis de píxeles anómalos no reveló la existencia de píxeles singulares que permitieran la distinción entre especies, debido, posiblemente, al gran tamaño de los píxeles con respecto al tamaño de las partículas singulares. El alcance de mayores resoluciones con equipos hiperespectrales en línea que posibiliten encontrar partículas singulares es una de las líneas futuras de investigación [Esta conclusión queda reflejada en la publicación: “Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99.].

11. Se propuso una metodología basada en el análisis textural de imágenes hiperespectrales. El uso de árboles de regresión permitió integrar la información textural y espectral en un mismo modelo que presentó mejores resultados que los modelos basados en información espectral solamente (92% versus 83%) [Esta conclusión queda reflejada en la publicación: “Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99].

12. Son necesarios estudios futuros con mayores y más variados conjuntos de muestras, así como la evaluación de modelos más sencillos basados en bandas específicas o ratios que permitan aumentar la rapidez del procesado [Esta conclusión reflejada en la publicación: “Using spectral and textural data extracted from hyperspectral near infrared spectroscopy imaging to discriminate between processed pork, poultry and fish proteins”, *Chemometrics and Intelligent Laboratory Systems*, 172 (2018) 90-99]

